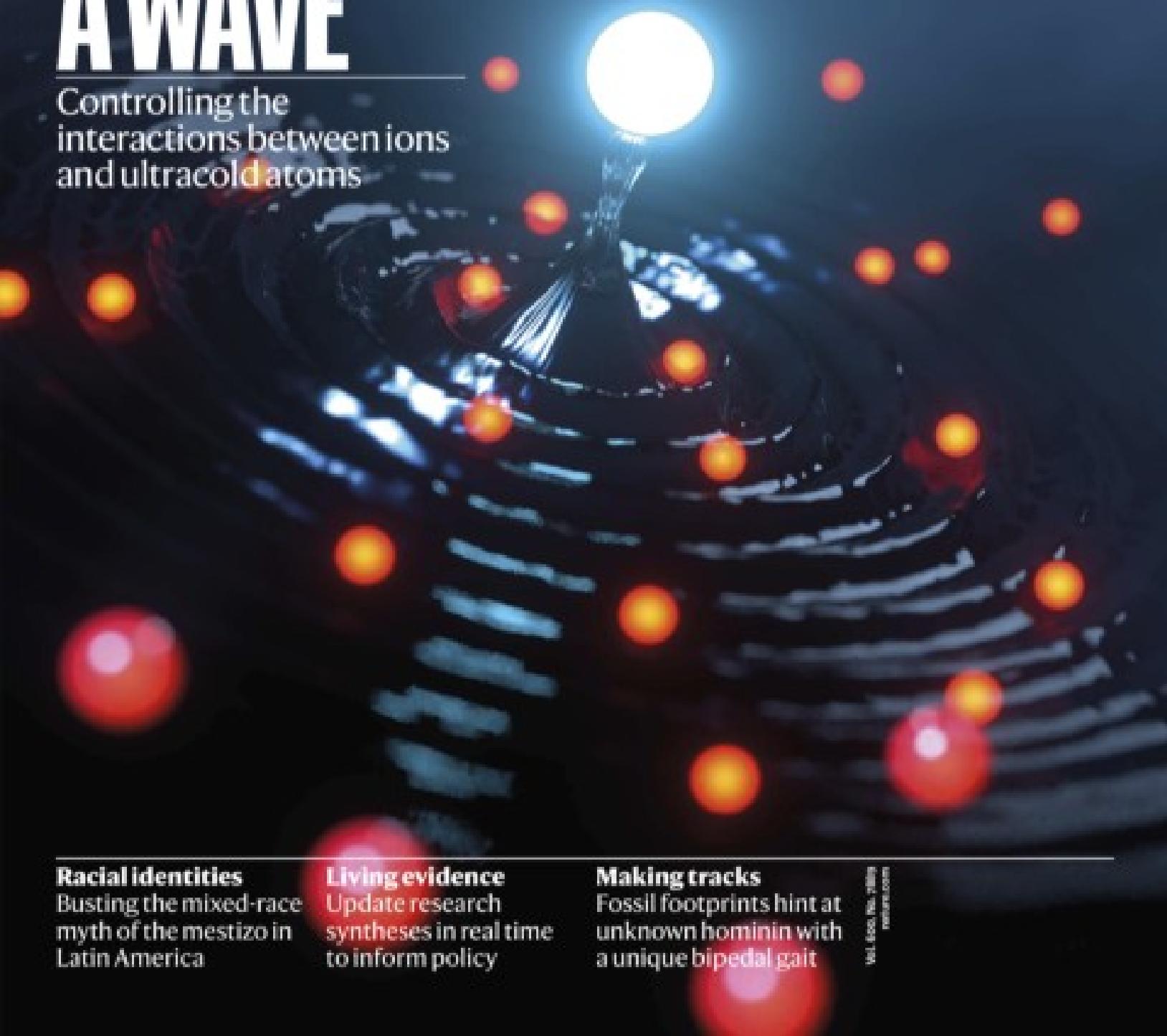


# nature

## CATCHING A WAVE

Controlling the interactions between ions and ultracold atoms



### Racial identities

Busting the mixed-race myth of the mestizo in Latin America

### Living evidence

Update research syntheses in real time to inform policy

### Making tracks

Fossil footprints hint at unknown hominin with a unique bipedal gait

# Nature.2021.12.18

[Sat, 18 Dec 2021]

- [This Week](#)
- [News in Focus](#)
- [Books & Arts](#)
- [Opinion](#)
- [Work](#)
- [Research](#)
- [Amendments & Corrections](#)

# This Week

- **[Replicating scientific results is tough — but essential](#)** [ 15 December 2021]  
Editorial • A high-profile replication study in cancer biology has obtained disappointing results. Scientists must redouble their efforts to find out why.
- **[Gender balance at Nature Conferences: an update](#)** [ 14 December 2021]  
Editorial • Nature has made progress in improving the representation and participation of women at scientific conferences — but there is still much more to do.
- **[India — stop looking down on international collaborations](#)**  
[ 14 December 2021]  
World View • Publications co-authored by scientists from the West are sometimes discounted by grant and hiring committees. That needs to change.
- **[Surging plastic use is fed by coal power — with deadly results](#)** [ 02 December 2021]  
Research Highlight • Coal-fired power plants are satisfying the world's voracious appetite for plastic, but their emissions are causing a wave of disease and deaths.
- **[Nervous stomach: lab-grown organs clench like the real thing](#)** [ 09 December 2021]  
Research Highlight • The most complex stomach organoids yet are made with cells from three layers found in developing embryos.
- **[Why cannabis reeks of skunk](#)** [ 09 December 2021]  
Research Highlight • Scientists discover sulfurous compounds that help to give pot a rank odour, holding out the promise of fragrance-free bud.
- **[Snake escape: imported reptiles gobble an island's lizards](#)**  
[ 09 December 2021]  
Research Highlight • King snakes brought to Gran Canaria have multiplied, wreaking havoc on all of its native reptiles.
- **[‘Sky river’ brought Iran deadly floods — but also welcome water](#)** [ 08 December 2021]  
Research Highlight • Moisture from the Mediterranean and Red seas helped to recharge an Iranian lake whose area has shrunk by 90% in recent decades.

- **It's alive! Bio-bricks can signal to others of their kind** [ 07 December 2021]  
Research Highlight • Large structures can be built from a microbial material with self-healing properties.
- **Running of the bulls tramples the laws of crowd dynamics** [ 09 December 2021]  
Research Highlight • The physics of crowds fails to describe behaviour of pedestrians at a famed bull-running festival.

---

| [Next section](#) | [Main menu](#) |

- EDITORIAL
- 15 December 2021
- Correction [16 December 2021](#)

# Replicating scientific results is tough — but essential

A high-profile replication study in cancer biology has obtained disappointing results. Scientists must redouble their efforts to find out why.



Funders and publishers need to take replication studies much more seriously than they do at present. Credit: Anne-Christine Poujoulat/AFP/Getty

Replicability — the ability to obtain the same result when an experiment is repeated — is foundational to science. But in many research fields it has proved difficult to achieve. An important and much-anticipated brace of

research papers now show just how complicated, time-consuming and difficult it can be to [conduct and interpret replication studies in cancer biology](#)<sup>1,2</sup>.

Nearly a decade ago, research teams organized by the non-profit Center for Open Science in Charlottesville, Virginia, and ScienceExchange, a research-services company based in Palo Alto, California, set out to systematically test whether selected experiments in highly cited papers published in prestigious scientific journals could be replicated. The effort was part of the high-profile Reproducibility Project: Cancer Biology (RPCB) initiative. The researchers assessed experimental outcomes or ‘effects’ by seven metrics, five of which could apply to numerical results. Overall, 46% of these replications were successful by three or more of these metrics, such as whether results fell within the confidence interval predicted by the experiment or retained statistical significance.

The project was launched in the wake of reports from drug companies that they could not replicate findings in many cancer-biology papers. But those reports did not identify the papers, nor the criteria for replication. The RPCB was conceived to bring research rigour to such retrospective replication studies.

## **Initial findings**

One of the clearest findings was that the effects of an experimental treatment — such as killing cancer cells or shrinking tumours — were drastically smaller in replications, overall 85% smaller, than what had been reported originally. It’s hard to know why. There could have been statistical fluke, for example; bias in the original study or in the replication; or lack of know-how by the replicators that caused the repeated study to miss some essential quality of the original.



### Half of top cancer studies fail high-profile reproducibility effort

The project also took more than five years longer than expected, and, despite taking the extra time, the teams were able to assess experiments in only one-quarter of the experiments they had originally planned to cover. This underscores the fact that such assessments take much more time and effort than expected.

The RPCB studies were budgeted to cost US\$1.3 million over three years. That was increased to \$1.5 million, not including the costs of personnel or project administration.

None of the 53 papers selected contained enough detail for the researchers to repeat the experiments. So the replicators had to contact authors for information, such as how many cells were injected, by what route, or the exact reagent used. Often, these were details that even the authors could not provide because the information had not been recorded or laboratory members had moved on. And one-third of authors either refused requests for more information or did not respond. For 136 of the 193 experimental effects assessed, replicators also had to request a key reagent from the original authors (such as a cell line, plasmid or model organism) because they could not buy it or get it from a repository. Some 69% of the authors were willing to share their reagents.

## Openness and precision

Since the reproducibility project began, several efforts have encouraged authors to share more-precise methodological details of their studies. *Nature*, along with other journals, introduced a reproducibility checklist in 2013. It requires that authors report key experimental data, such as the strain, age and sex of animals used. Authors are also encouraged to deposit their experimental protocols in repositories, so that other researchers can access them.



### [Understand the real reasons reproducibility reform fails](#)

Furthermore, the ‘Landis 4’ criteria were published in 2012 to promote rigorous animal research. They include the requirement for blinding, randomization and statistically assessed sample sizes. Registered Reports, an article format in which researchers publish the design of their studies before doing their experiments, is another key development. It means that [‘null effects’ are more likely to be published than buried in a file drawer](#). The project team found that null effects were more likely to be replicated; 80% of such studies passed by three metrics, compared with only 40% of ‘positive effects’.

Harder to resolve is the fact that what works in one lab might not work in another, possibly because of inherent variation or unrecognized methodological differences. Take the following example: one study tracked whether a certain type of cell contributes to blood supply in tumours<sup>3</sup>. Tracking these cells required that they express a ‘reporter’ molecule (in this case, green fluorescent protein). But, despite many attempts and tweaks, the replicating team couldn’t make the reporter sufficiently active in the cells to be tracked<sup>4</sup>, so the replication attempt was stopped.

The RPCB teams vetted replication protocols with the original authors, and also had them peer reviewed. But detailed advance agreement on experimental designs will not necessarily, on its own, account for setbacks encountered when studies are repeated — in some cases, many years after the originals. That is why another approach to replication is used by the US Defense Advanced Research Projects Agency (DARPA). In one DARPA programme, research teams are assigned independent verification teams. The research teams must help to troubleshoot and provide support for the verification teams so that key results can be obtained in another lab even before work is published. This approach is built into programme requirements: 3–8% of funds allocated for research programmes go towards such verification efforts<sup>5</sup>.

Such studies also show that researchers, research funders and publishers must take replication studies much more seriously. Researchers need to engage in such actions, funders must ramp up investments in these studies, and publishers, too, must play their part so that researchers can be confident that this work is important. It is laudable that the press conference announcing the project’s results included remarks and praise by the leaders of the US National Academies of Sciences, Engineering, and Medicine and the National Institutes of Health. But the project was funded by a philanthropic investment fund, Arnold Ventures in Houston, Texas.

The entire scientific community must recognize that replication is not for replication’s sake, but to gain an assurance central to the progress of science: that an observation or result is sturdy enough to spur future work. The next wave of replication efforts should be aimed at making this everyday essential easier to achieve.

*Nature* **600**, 359–360 (2021)

doi: <https://doi.org/10.1038/d41586-021-03736-4>

## Updates & Corrections

- **Correction 16 December 2021:** This article originally mischaracterized the RPCB's analysis of replication attempts. Rather than recording seven experimental outcomes, it assessed experimental effects using seven metrics, and it also assessed 193 experimental effects not 193 experiments.

## References

1. 1.  
Errington, T. M., Denis, A., Perfito, N., Iorns, E. & Nosek, B. A. *eLife* **10**, e67995 (2021).
2. 2.  
Errington, T. M. *et al.* *eLife* **10**, e71601 (2021).
3. 3.  
Ricci-Vitiani, L. *et al.* *Nature* **468**, 824–828 (2010).
4. 4.  
Errington, T. M. *et al.* *eLife* **10**, e73430 (2021).
5. 5.  
Raphael, M. P., Sheehan, P. E. & Vora, G. J. *Nature* **579**, 190–192 (2020).

| [Section menu](#) | [Main menu](#) |

- EDITORIAL
- 14 December 2021

# Gender balance at Nature Conferences: an update

*Nature* has made progress in improving the representation and participation of women at scientific conferences — but there is still much more to do.

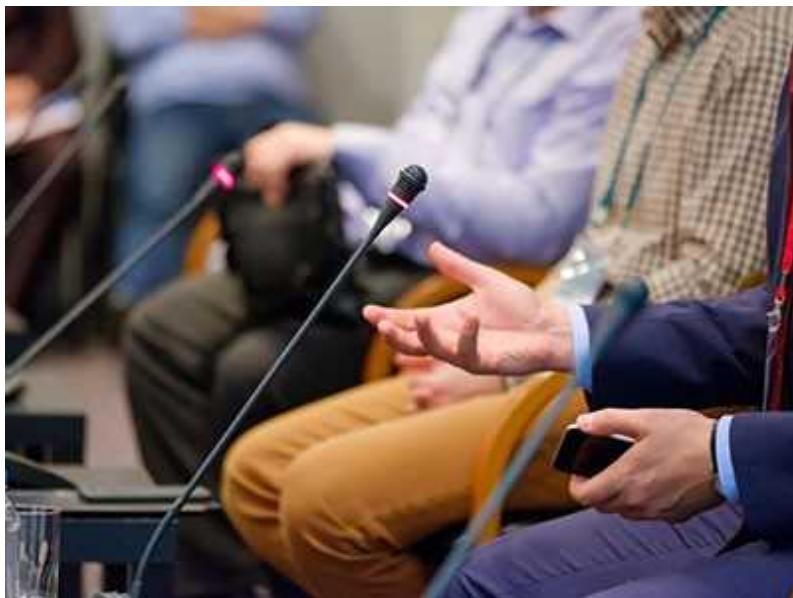


Studies are starting to show that more women participate in conference Q&As when session chairs invite women to ask opening questions after a talk. Credit: Kena Betancur/Getty

At the end of 2019, *Nature* pledged to work harder to help to address the entrenched gender inequity at scientific conferences ([\*Nature\* 576, 182; 2019](#)). We looked closely at gender diversity at Nature Conferences

(conferences curated by editors of the Nature Portfolio journals), and what we saw was simply not good enough. We introduced a code of conduct, including pledges to have no all-male panels and to invite an equal percentage of women (including all those who identify as women) and men as speakers at all our conferences.

Two years on and these decisions have yielded results. Women comprised 29% of keynote speakers at Nature Conferences between 2016 and 2018 (15 out of 51 speakers across 27 events). That number increased to 48% (14 out of 29) in the 20 meetings that took place in 2020 and 2021.



### Nature Conferences: no more ‘manels’

During the same period, women represented 43% of conference organizing committees (59 out of 130) and 51% of speakers for short talks (33 of 65). It’s an encouraging start, but the trends cannot stop there. In 2021, all Nature Conferences were virtual, like almost all other scientific events. It’s early days, but initial reports suggest that online formats can be more inclusive than in-person events ([S. Sarabipour \*eLife\* 9, e62668; 2020](#)). It is absolutely essential that these modest gains are not reversed once in-person events return.

Best-practice guidance is being refined all the time. The advocacy group 500 Women Scientists is working with several major scientific organizations,

including the Aspen Global Change Institute, the American Geophysical Union, Colorado State University, the Earth Science Women's Network, Georgia Sea Grant and the team behind the virtual seminar series Pal(a)eoPERCS, to update their inclusive scientific meetings guide (see [go.nature.com/3ilz3e5](https://go.nature.com/3ilz3e5)). The guide aims to share good practice, including tools to help ensure that events are more inclusive. Nature Conferences will strive to use these tools, and we hope that other conference organizers will too.

Data from the UK Society for Endocrinology's annual national conference are the latest to show that even when meetings have roughly equal numbers of male and female delegates, women attendees participate less and tend to ask fewer and shorter questions ([V. Salem et al. \*Lancet Diabetes Endocrinol.\* 9, 556–559; 2021](https://doi.org/10.1016/j.lindia.2021.03.003)). A simple intervention improved things: when conference organizers were asked to have more female session chairs, and those chairs were asked to invite opening questions from women wherever possible, questions from women rose to 35%, from 24% the previous year.



### How to banish manels and manferences from scientific meetings

“If women are not visible at conferences, they cannot act as role models for junior academics, creating a self-perpetuating cycle,” the paper’s authors point out. Nature’s insistence on gender parity, where possible, at its

conferences is just one step towards encouraging more women to take part in their communities' discussions.

Codes of conduct for scientific meetings, such as that introduced by Nature Conferences and others before us, are crucial. They are becoming the norm, with many large learned/academic societies making commitments to improve diversity across their activities — and there are encouraging signs that they are starting to pay off.

Codes of conduct need to be enforced, and there need to be checks in place to ensure that they are followed and that they continue to be effective at truly improving diversity — not just in attendance, but in participation, too. We also acknowledge that our efforts need to be broader, and not focused solely on gender.

There's a very long way to go to achieve full equity, inclusion and diversity at scientific conferences. The Nature journals are committed to achieving inclusivity for the good of science and society. We are proud to have made a small change and understand and accept that there is much more that we can and must do.

*Nature* **600**, 360 (2021)

doi: <https://doi.org/10.1038/d41586-021-03735-5>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03735-5>

- WORLD VIEW
- 14 December 2021

# India — stop looking down on international collaborations



Publications co-authored by scientists from the West are sometimes discounted by grant and hiring committees. That needs to change.

- [Arun Kumar Shukla](#) ✉

During my first week at my new job in 2014, a senior colleague called me into his office to share experience and wisdom. He advised that if a research publication from my laboratory had foreign co-authors, then it would count little towards my career progression.

I was taken aback. I was starting the position I had always aspired to — as an independent faculty member in my home country — but I certainly wanted to maintain the connections I had established during my studies

overseas and to make new global ones. Having trained with three Nobel laureates, I knew the power of collaboration to drive cutting-edge science.

I chalked up this advice as one person's idiosyncratic opinion, and continued to nurture a broad network of international collaborators. This was crucial for my study of G-protein-coupled receptors, which pass diverse messages across the cell or nuclear surface and are the targets of powerful drugs, including those prescribed to treat heart failure, hypertension, allergies and mental illnesses.

Today, the research in my lab is built on extensive connections across the globe. It has been recognized with several national awards, which I think reflects a growing appreciation of overseas collaborations. Dozens of students and early-career researchers whom I mentor see how working with scientists in other nations helps them to achieve greater insights and holistic understanding. I am incredibly proud of them.



### Reviewers are blinkered by bibliometrics

Alas, I still often hear that the committees that evaluate promotions and awards look down on work that includes out-of-country collaborators. This idea has come up in casual conversations. Some senior scientists consider it essential advice, even writing it in e-mails. Their contention? If there is a foreign author on a paper, even in the middle of the list, then the paper has

been published in a high-visibility outlet because of that author and not because of the science.

I don't want to cast assessment committees in a bad light. My aim is to raise an awareness of this bias, which might sometimes be unconscious. It runs counter to national strategy.

In the past few years, India has invested heavily in huge science projects and international partnerships with institutions such as Wellcome, the London-based research funder, and the funding agency UK Research and Innovation, and the impact is starting to show. For example, the Mega Science Vision from the Indian government explicitly supports and promotes international collaborations. The Department of Science and Technology, a major funding agency in the country, has launched several bi- and multilateral collaborative funding schemes over the past few years. India has also joined with international research-funding bodies, such as the Human Frontier Science Program and the European Molecular Biology Organization.

But junior colleagues still sometimes suffer because of this bias. One told me about an astonishing experience at an interview for a faculty position. A senior member of the hiring committee said he should not even have been invited because his research papers included foreign authors, and therefore it was too difficult to assess his contribution. This was a senior scientist entrusted with recruiting scientific leaders of the next generation.

If it is possible to gauge the contribution of applicants from international institutes, then it is possible to evaluate a local or internal candidate's work with foreign collaborators, especially with so many journals including an author-contributions section. This young, enthusiastic researcher was holding back tears. I felt that the system had failed him.



## Farewell authors, hello contributors

In surveys during 2013 and 2015, taken by thousands of scientists from the United States, the United Kingdom, India, Italy, Taiwan, Hong Kong, Turkey and France, Indian scientists were the most likely to say that none of their scientific collaborators over the previous year had been from other countries ([K. R. W. Matthews \*et al. Account. Res.\* 27, 477–495; 2020](#)). Yet papers with international authorship [tend to be more highly cited and impactful](#).

I have personally seen the power of international collaborations to expand the quality of my research in India. Most labs and departments — even those in leading Indian research institutions — are relatively small, limiting the number of people available to work together on a project. For example, many crystallographers in foreign countries can simply drive to a synchrotron to collect X-ray diffraction data; in India, we have to plan several months in advance to send crystals and collect data at beamlines abroad. Moreover, one lab cannot acquire expertise in all possible methodologies to deal with a question, and sometimes technologies are available abroad before they can be accessed here: cryo-electron microscopy (cryo-EM) is an apt example.

In 2018, a US\$8.6-million (650-million-rupee) national cryo-EM facility opened in Bengaluru, after a cross-country lobbying effort. It was followed

by four more such facilities, approved last year with a total cost of about \$16 million. Now, protein structures can be studied using this powerful technique in the country. Still, collaborators across the world are crucial for support, including, for example, matching funds from Wellcome in an initiative called the India Alliance.

Anti-collaboration sentiment must end. It undercuts scientists' willingness to engage across borders. This limits their own work, and hampers the progress of science in India and globally.

*Nature* **600**, 361 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-03700-2>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03700-2>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 02 December 2021

# Surging plastic use is fed by coal power — with deadly results

Coal-fired power plants are satisfying the world's voracious appetite for plastic, but their emissions are causing a wave of disease and deaths.



The growing demand for plastic is being fed by coal-fired plastic production, especially in Asian countries. Credit: Jeffrey Greenberg/Universal Images Group/Getty

Plastic production has soared in countries that rely heavily on coal for energy, leading to a rise in carbon emissions and particle pollution — which, in turn, has led to a steep increase in illness and death<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 362 (2021)

doi: <https://doi.org/10.1038/d41586-021-03613-0>

## **References**

1. 1.

Cabernard, L., Pfister, S., Oberschelp, C. & Hellweg, S. *Nature Sustain.* <https://doi.org/10.1038/s41893-021-00807-2> (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03613-0>

- RESEARCH HIGHLIGHT
- 09 December 2021

# Nervous stomach: lab-grown organs clench like the real thing

The most complex stomach organoids yet are made with cells from three layers found in developing embryos.

 Human gastric tissue engineered from pluripotent stem cells.

A replica stomach grown in the laboratory includes muscle (red), neurons (green) that control stomach-tissue contraction and tissue (white) containing acid glands. Credit: Alexandra Eicher

Scientists have created a sophisticated mini-stomach, complete with nerves that coordinate muscle contraction and glands that make acid, providing hope that such ‘organoids’ could one day be used to repair damaged gastrointestinal tracts<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.  
VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 362 (2021)

doi: <https://doi.org/10.1038/d41586-021-03648-3>

## **References**

1. 1.

Eicher, A. K. *et al.* *Cell Stem Cell*  
<https://doi.org/10.1016/j.stem.2021.10.010> (2021).

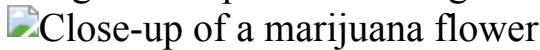
---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03648-3>

- RESEARCH HIGHLIGHT
- 09 December 2021

# Why cannabis reeks of skunk

Scientists discover sulfurous compounds that help to give pot a rank odour, holding out the promise of fragrance-free bud.



Sulfur-laden compounds in *Cannabis sativa* give the plant a musky smell that is both savoured and loathed. Credit: Pablo Porciuncula  
Brune/AFP/Getty

Add cannabis to the list of pungent plants that get their strong smell from sulfur<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 362 (2021)

*doi: <https://doi.org/10.1038/d41586-021-03650-9>*

## **References**

1. 1.

Oswald, I. W. H. *et al.* *ACS Omega* **6**, 31667–31676 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03650-9>

- RESEARCH HIGHLIGHT
- 09 December 2021

# Snake escape: imported reptiles gobble an island's lizards

King snakes brought to Gran Canaria have multiplied, wreaking havoc on all of its native reptiles.



The Gran Canaria giant lizard is large and feisty — but it is no match for the California kingsnakes now rampaging through the lizard's habitat.

Credit: Nick Upton/Nature Picture Library

Two of the three native reptiles on to Gran Canaria have nearly vanished from some parts of the Spanish island — eaten by an invasive snake species originally imported as a pet<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 362 (2021)

doi: <https://doi.org/10.1038/d41586-021-03647-4>

## **References**

1. 1.

Piquet, J. C. & López-Darias, M. *Proc. R. Soc. B*  
<https://doi.org/10.1098/rspb.2021.1939> (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03647-4>

- RESEARCH HIGHLIGHT
- 08 December 2021

## ‘Sky river’ brought Iran deadly floods — but also welcome water

Moisture from the Mediterranean and Red seas helped to recharge an Iranian lake whose area has shrunk by 90% in recent decades.

 Colourful plastic swans sit on an abandoned beach by Lake Urmia.

Lake Urmia was an Iranian tourist centre until greatly diminished by human water use. It was partially replenished by an atmospheric river in 2017.  
Credit: Dominika Zarzycka/NurPhoto/Getty

Devastating floods that struck Iran in 2017 were caused by a ‘sky river’ that ferried in water from hundreds or thousands of kilometres away — and that brought benefits, as well as destruction<sup>1</sup>.

### Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 363 (2021)

doi: <https://doi.org/10.1038/d41586-021-03646-5>

## **References**

1. 1.

Dezfuli, A., Bosilovich, M. G. & Barahona, D. *Geophys. Res. Lett.* <https://doi.org/10.1029/2021GL095441> (2021).

2. 2.

Dezfuli, A. *Bull. Am. Meteorol. Soc.* **101**, E394–E400 (2020).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03646-5>

- RESEARCH HIGHLIGHT
- 07 December 2021

## It's alive! Bio-bricks can signal to others of their kind

Large structures can be built from a microbial material with self-healing properties.



Bricks made of living organisms can be joined into a free-standing arch.  
Credit: R. M. McBee *et al.*/*Nature Mater.*

‘Living’ bricks made from bacteria and fungi can heal themselves and send signals to each other while serving as building blocks for sturdy structures<sup>1</sup>.

### Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 363 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-03644-7>

## **References**

1. 1.

McBee, R. M. *et al.* *Nature Mater.* <https://doi.org/10.1038/s41563-021-01123-y> (2021).

---

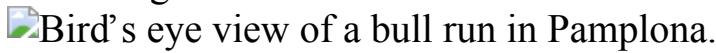
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03644-7>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 09 December 2021

# Running of the bulls tramples the laws of crowd dynamics

The physics of crowds fails to describe behaviour of pedestrians at a famed bull-running festival.



The bull-running festival in Pamplona, Spain. Analysis shows that the clump of people just ahead of a charging bull gains both speed and density, a pattern contrary to that seen in other crowds. Credit: Ander Gillenea/AFP/Getty

Traditional bull runs — whose participants try to outrun animals charging down a street — defy the dynamics commonly seen in the motion of crowds, researchers have found<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.  
VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 363 (2021)

doi: <https://doi.org/10.1038/d41586-021-03645-6>

## **References**

1. 1.

Parisi, D. R. *et al.* *Proc. Natl Acad. Sci. USA* **118**, e2107827118 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03645-6>

# News in Focus

- **[Merck's COVID pill and rare ancient skull](#)** [ 15 December 2021]  
News Round-Up • The latest science news, in brief.
- **[Omicron likely to weaken COVID vaccine protection](#)** [ 08 December 2021]  
News • Early lab results suggest that existing vaccines could be less effective against the fast-spreading coronavirus variant, but boosters should improve immunity.
- **[Half of top cancer studies fail high-profile reproducibility effort](#)** [ 09 December 2021]  
News • Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.
- **[What Sci-Hub's latest court battle means for research](#)** [ 13 December 2021]  
News • Delhi court will scrutinize whether the pirate paper website falls foul of India's copyright law. The verdict could have implications for academic publishers further afield.
- **[DeepMind AI tackles one of chemistry's most valuable techniques](#)** [ 10 December 2021]  
News • Machine-learning algorithm predicts material properties using electron density.
- **[Science misinformation alarms Francis Collins as he leaves top NIH job](#)** [ 03 December 2021]  
News Q&A • The genome project leader reflects on his 12 years at the helm of a juggernaut biomedical agency, and what lies ahead.
- **[How the mixed-race mestizo myth warped science in Latin America](#)** [ 13 December 2021]  
News Feature • Researchers are trying to dismantle the flawed concept of homogeneous racial mixing that has fostered discrimination in Mexico, Brazil and other countries.
- **[The race to make vaccines for a dangerous respiratory virus](#)** [ 10 December 2021]  
News Feature • Millions of people a year are hospitalized by respiratory syncytial virus and tens of thousands die. After decades of failure, four vaccines are now in late-stage trials.

---

| [Next section](#) | [Main menu](#) | [Previous section](#) |

- NEWS ROUND-UP
- 15 December 2021

## Merck's COVID pill and rare ancient skull

The latest science news, in brief.



Leti's skull was found in Rising Star Cave. Credit: Wits University

### Rare ancient hominin skull discovered in South Africa

The first partial skull of a 4- to 6-year-old *Homo naledi* child, who died almost 250,000 years ago, has been [found in the depths of Rising Star Cave](#)

near Johannesburg, South Africa, where the ancient hominin species was first discovered in 2015.

The skull comprises 28 fragments and 6 teeth that researchers have pieced together into this reconstruction ([J. K. Brophy et al. \*Palaeoanthropology\* <http://doi.org/g46n>; 2021](#)). The team named the skull's owner Leti, after the Setswana word *letimela*, which means 'the lost one'. It is a rare find — juvenile hominin remains are usually thin and extremely fragile.

## Merck's COVID pill loses its lustre

Things are looking less than rosy for molnupiravir, one of two antiviral pills that have caused excitement because preliminary clinical-trial results showed that they can significantly reduce hospitalizations and deaths from COVID-19. A US Food and Drug Administration (FDA) advisory committee that met on 30 November voted only narrowly to [recommend that the drug candidate receive emergency authorization](#). As *Nature* went to press, the FDA itself had yet to issue a decision.

The agency's lengthy deliberations could signal uncertainties about the antiviral's benefits: full trial data submitted to the FDA suggest that molnupiravir is less effective than originally thought, dampening scientists' hopes that the relatively cheap and easy-to-administer treatment might change the course of the pandemic.

The results showed that the antiviral, which was developed by the pharmaceutical firm Merck, based in Kenilworth, New Jersey, and the biotechnology company Ridgeback Biotherapeutics in Miami, Florida, decreased the risk of hospitalization from COVID-19 by 30% — down from a 50% reduction observed early in the trial. "That's not all that good," says Katherine Soley-Radtke, a medicinal chemist at the University of Maryland, Baltimore County. Monoclonal antibody treatments reduce the risk of severe COVID-19 by up to 85%, although they are costly and need to be administered intravenously.

Nicholas Kartsonis, the senior vice-president of clinical research at Merck, told the FDA advisory committee that the company couldn't explain the

starkly different results, which have not been peer reviewed.

*Nature* **600**, 365 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-03701-1>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03701-1>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 08 December 2021

# Omicron likely to weaken COVID vaccine protection

Early lab results suggest that existing vaccines could be less effective against the fast-spreading coronavirus variant, but boosters should improve immunity.

- [Ewen Callaway](#)



A passenger at Schipol Airport in Amsterdam is tested for COVID-19 from a flight from South Africa after disembarking a flight from South Africa,

where the Omicron SARS-CoV-2 variant was first reported. Credit: Pierre Crom/Getty

The fast-spreading Omicron SARS-CoV-2 variant is highly likely to compromise some of the protection from vaccines, suggest the first laboratory studies of Omicron's ability to evade immunity.

But the preliminary results — released overnight by teams in South Africa, Germany and Sweden, as well as by the Pfizer–BioNTech collaboration — hint that protection conferred by existing COVID-19 vaccines won't be totally wiped out, and that boosters should improve immunity to Omicron.



### How bad is Omicron? What scientists know so far

“We’re likely to see reduced effectiveness of vaccines against preventing infection,” says Penny Moore, a virologist at the University of Witwatersrand in Johannesburg, South Africa, who co-authored one of the studies. “I think it’s a strong argument to get boosters out there.”

The studies, which measure the capacity of antibodies in people’s blood to block the infection of cells in a dish, have not yet been peer reviewed, and do not tell researchers the extent to which vaccines’ ability to protect against COVID-19 — in particular, its most severe forms — could be compromised by Omicron.

“We still need to wait for more effectiveness data and clear signals from the places where this is blowing up first,” says Ben Murrell, an interdisciplinary virologist and immunologist at the Karolinska Institute in Stockholm, who co-led one of the studies.

## Many mutations

Researchers in Botswana and South Africa [identified Omicron in late November](#), and teams worldwide have since been racing to understand the variant’s properties and the risks that it poses. Preliminary data from South Africa and elsewhere suggest that the variant is highly transmissible — spreading several times faster than Delta — and might be able to infect people who are immune to other variants.

Omicron carries a large number of mutations in its spike protein — the prime target of immune responses — and some of these changes, when present in other variants, affect the ability of antibodies to recognize the virus and block infection.

Scientists used two types of laboratory assay to test how well Omicron can evade neutralizing, or virus-blocking, antibodies triggered by vaccination and infection. One approach uses infectious SARS-CoV-2 particles, typically isolated from individuals infected with Omicron. The other relies on pseudovirus particles — genetically modified versions of another virus (often HIV) that use the SARS-CoV-2 spike protein to infect cells.



### Omicron-variant border bans ignore the evidence, say scientists

The results from the four separate teams all suggest that Omicron blunts the potency of neutralizing antibodies more extensively than any other circulating SARS-CoV-2 variant. But the magnitude of Omicron's impact varied between the studies, which examined blood from people with different vaccination and infection histories.

A study led by virologist Alex Sigal, at the Africa Health Research Institute in Durban, South Africa, found that serum — the antibody-containing portion of blood — from 12 people who received the Pfizer–BioNTech vaccine was around 40 times less potent against Omicron, on average, than against an earlier strain of SARS-CoV-2. That finding was similar to the results from two other studies: one reported by Pfizer and BioNTech in an 8 December press release, and the other released on Twitter and later posted on medRxiv<sup>1</sup> by virologist Sandra Ciesek at the Goethe University Frankfurt, Germany.

A fourth study, led by Murrell and virologist Daniel Sheward, also at the Karolinska Institute, reported a smaller reduction in levels of Omicron neutralizing antibodies in two groups of participants: 17 health-care workers, who had all been previously infected, and 17 Swedish blood donors. The researchers cannot determine the vaccine status of the anonymous blood

donors, but say they will soon update their paper with vaccination information from the health-care workers.

Despite differences in the labs' results — which are common in such virus neutralization assays — the conclusions are similar, and show that Omicron's effects on neutralizing antibodies are “not complete knockouts”, says Murrell. “The magnitude is still a little up for question.”

## Booster protection

The results suggest that vaccines' effectiveness is likely to be significantly modified by Omicron — but precisely how much is hard to say. Sigal's team found that people who had already been infected before vaccination tended to have higher levels of neutralizing antibodies against Omicron than vaccinated people with no known history of infection. “I think retaining some neutralization against Omicron can only be helpful,” says Moore, a co-author on the study, whose lab is also working on neutralization experiments.



### Omicron is supercharging the COVID vaccine booster debate

A previous case of COVID-19 isn't the only way to improve antibody levels against Omicron. The Pfizer–BioNTech study found that people who had

received a third dose of its vaccine had neutralizing antibody levels against Omicron comparable to those against other SARS-CoV-2 variants that were triggered by 2 vaccine doses. On the basis of those results, “we expect significant protection against any type of COVID-19 mediated by Omicron in individuals who have received the third vaccine”, said BioNTech’s chief executive, Uğur Şahin, at a press conference on 8 December.

Danny Altmann, an immunologist at Imperial College London, agrees that jacking up antibody levels with booster shots should help protect against Omicron, just as boosters have improved protection against the Delta variant. “Omicron is scarier than anything we’ve known before, because it’s a little bit worse still than Delta. But we were in quite a bad situation with Delta in unboosted populations,” Altmann says.

Jesse Bloom, an evolutionary biologist at the Fred Hutchinson Cancer Research Center in Seattle, Washington, says that it will be important to determine the extent to which immune mechanisms other than neutralizing antibodies, such as T cells, ameliorate severe disease caused by infection.

It will also be important to see further studies confirming the latest results, because variables such as the type of cell used can affect conclusions, says Pei-Yong Shi, a virologist at the University of Texas Medical Branch at Galveston. “In the next week or ten days, there will be a lot of confirmatory results coming out,” he says.

*Nature* **600**, 367-368 (2021)

doi: <https://doi.org/10.1038/d41586-021-03672-3>

## References

1. 1.

Wilhelm, A. *et al.* Preprint at medRxiv  
<https://doi.org/10.1101/2021.12.07.21267432> (2021).

| [Section menu](#) | [Main menu](#) |

- NEWS
- 09 December 2021

# Half of top cancer studies fail high-profile reproducibility effort

Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.

- [Asher Mullard](#)



Vague experimental protocols was one barrier to replication that researchers encountered.Credit: Patrick Hertzog/AFP/Getty

A US\$2-million, 8-year attempt to replicate influential preclinical cancer research papers has released its final — and disquieting — results. Fewer than half of the experiments assessed stood up to scrutiny, reports the Reproducibility Project: Cancer Biology (RPCB) team in *eLife*<sup>1,2</sup>. The project — one of the most robust reproducibility studies performed so far — documented how hurdles including vague research protocols and uncooperative authors delayed the initiative by five years and halved its scope.

“These results aren’t surprising. And, simultaneously, they’re shocking,” says Brian Nosek, an RPCB investigator and executive director of the Center for Open Science in Charlottesville, Virginia. Although initially planning to repeat 193 experiments from 53 papers, the team ran just 50 experiments from 23 papers.



### Reproducibility: expect less of the scientific paper

The low replication rate is “frankly, outrageous”, says Glenn Begley, an oncologist and co-founder of Parthenon Therapeutics in Cambridge, Massachusetts, who was not involved in the study. But it isn’t unexpected, he agrees. In 2012, while at the biotech firm Amgen in Thousand Oaks, California, Begley’s team helped to draw attention to growing evidence of a ‘reproducibility crisis’, the concern that many research findings cannot be replicated. Over the previous decade, his haematology and oncology team

had been able to confirm the results of only [6 of the 53 \(11%\) landmark papers it assessed](#), despite working alongside the papers' original authors.

Other analyses have reported low replication rates in drug discovery, [neuroscience](#) and [psychology](#).

## Double take

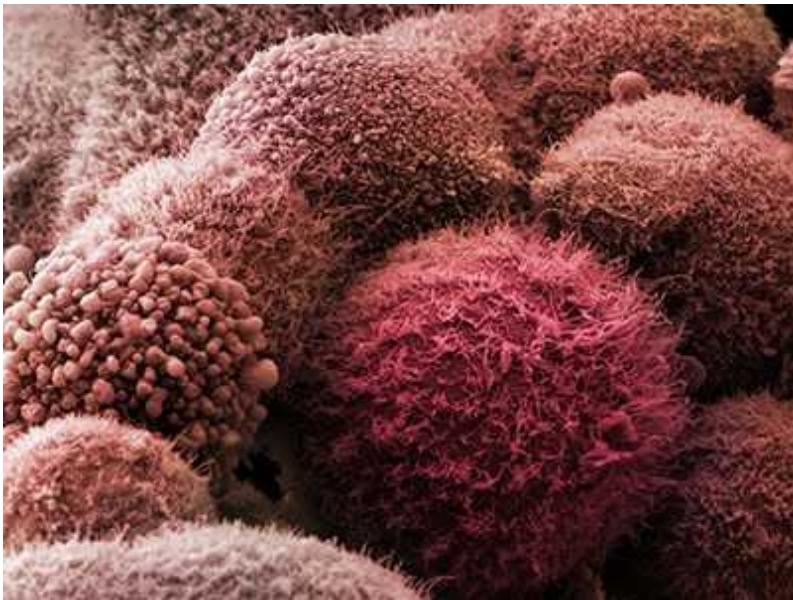
The RPCB — a partnership between the Center for Open Science and Science Exchange, a marketplace for research services in Palo Alto, California — launched in 2013. Funded by the philanthropic investment fund Arnold Ventures, headquartered in Houston, Texas, the collaborators set out to systematically reproduce experiments in 53 high-profile papers published during 2010–12 in journals including *Nature*, *Science* and *Cell*.

The project focused on preclinical cancer research because early hints at low reproducibility rates came from this space — animal studies, in particular, seemed difficult to reproduce. By selecting high-impact papers, the team focused on the research that most shapes the field.

The RPCB started publishing its findings in 2017, and these [hinted at the messy results to come](#). The researchers now summarize their overall findings in two papers published on 7 December.

The first of these papers<sup>1</sup> catalogues the hurdles the researchers encountered. For every experiment they set their sights on, for example, they needed to contact the authors for advice on experimental design because the original papers lacked data and details. They deemed 26% of authors “extremely helpful”, sometimes spending months tracking down answers and sharing reagents. But 32% were “not at all helpful” — often ignoring queries altogether.

“Everyone always talks about this problem. But here, we've actually got data on how prevalent it is,” says Manoj Lalu, a clinician-researcher who studies data reproducibility at the Ottawa Hospital Research Institute in Canada.



## Cancer reproducibility project releases first results

This lack of cooperation, alongside the need to modify or overhaul protocols once experiments were under way, took a toll. On average, the team needed 197 weeks to replicate a study. And as costs added up to \$53,000 per experiment — about twice what the team had initially allocated — the [project's budget couldn't cover its original ambition.](#)

The second study<sup>2</sup> delves into the overall results of these experiments in detail. By one analysis, only 46% of the attempted replications confirmed the original findings. And, on average, the researchers observed effect sizes that were 85% smaller than originally reported.

The experiments with the biggest effect sizes were those most likely to be replicated. Animal experiments fared worst, mainly because *in vivo* experiments tend to yield smaller effect sizes than do *in vitro* experiments.

## **Counterclaims**

Not everyone is convinced that the study has merit. Pushback came especially from researchers whose findings were not successfully replicated.

“I’m not sure there is much value in these one-shot experiments,” says Erkki Ruoslahti, a cancer biologist at the Sanford Burnham Prebys in La Jolla, California. In 2017, the RPCB team reported that it could not confirm a finding made by Ruoslahti’s team, but Ruoslahti counters that external laboratories have replicated the disputed result at least 20 times. A drug candidate resulting from this work is now in phase II trials. “It’s hard for me to believe that half of all papers out there would not be valid,” he says.

Dean Tang, a cancer biologist at the Roswell Park Comprehensive Cancer Center in Buffalo, New York, is also circumspect. The RPCB reported<sup>3</sup> in 2019 that it could not replicate some work from his lab. But, he argues, the replicators deviated from their experimental plan, relied on fewer and different cell lines from those used in the original study, and didn’t double-check their own work. “We believe all published work deserves at least a minimum of 3 attempted replicates before being discounted,” Tang and a colleague wrote in 2019 in response to the project’s findings.

But replication is extremely hard, says Olavo Amaral, a coordinator of the Brazilian Reproducibility Initiative and a neuroscientist at the Federal University of Rio de Janeiro, Brazil. “You can never do it exactly the same,” he says. Does it matter if you shake a tube up and down instead of side to side? How do you account for different baseline readings? Figuring out when and how to stay true to an experimental protocol is part of the emerging science of replication, he says.

### Understand the real reasons reproducibility reform fails

Failure to replicate alone is not necessarily cause for concern, says Nosek. Some preliminary findings are distractions, but contradictory follow-up results can lead to deeper scientific insights. The RPCB was not set up to call out or invalidate specific studies, adds Nosek. Replication, like science, is about the total body of evidence. Rather, he says, the goal was to capture a snap shot of the drivers and the magnitude of the reproducibility crisis, with an eye towards system-level solutions.

The real problem is the time, money and effort that are wasted in finding the signals amid the noise, says Tim Errington, the RPCB’s project leader and director of research at the Center for Open Science. “How well are we using

our resources? And how are we learning new knowledge? This is the place to keep pushing, across disciplines.”

## Culture shift

There is no shortage of proposed fixes: for example, *in vitro* and animal studies can benefit from blinding, bigger sample sizes, greater statistical rigour and preregistration of study plans. Papers should make fewer claims and provide more proof, researchers suggest. Data sharing and reporting requirements need to be baked into scientific processes.

But stakeholders also need to address the incentives and research cultures that stand in the way of replication, says Nosek. Researchers who have published high-profile papers have little to gain from participating in confirmatory analyses, he points out, and much to lose. Replication attempts are often seen as threats rather than as compliments or opportunities for progress, he says. “That kind of culture does not help this ethos of self-correction. We are really about changing the entire research culture,” says Nosek.

There is also currently little support for the researchers who show that something doesn’t work, or who focus on the causes of variability between labs, says Lalu. “Hopefully, this will provide some people sober second thought about how we’re going to approach this moving forward.”

Begley sees evidence that attitudes are already changing. “When I first presented my findings, people got very hostile. Now, people accept that there’s a problem, and ask about what needs to happen to change this.”

*Nature* **600**, 368-369 (2021)

doi: <https://doi.org/10.1038/d41586-021-03691-0>

## References

1. 1.

Errington, T. M., Denis, A., Perfito, N., Iorns, E. & Nosek, B. A. *eLife* **10**, e67995 (2021).

2. 2.

Errington, T. M. *et al.* *eLife* **10**, e71601 (2021).

3. 3.

Yan. X. *et al.* *eLife* **8**, e43511 (2019).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03691-0>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 13 December 2021
- Clarification [14 December 2021](#)

# What Sci-Hub's latest court battle means for research

Delhi court will scrutinize whether the pirate paper website falls foul of India's copyright law. The verdict could have implications for academic publishers further afield.

- [Holly Else](#)



Access to Sci-Hub has been blocked in several countries, but the site argues its activities are permitted under India's copyright law. Credit: Sharaf Maksumov/Shutterstock

Sci-Hub, the popular website that offers access to millions of pirated research papers and books, is no stranger to legal action. But, for the first time, the site is defending its operations in court, in a copyright case filed in India by a group of major publishers.

In a lawsuit presented in Delhi's high court, the American Chemical Society, Elsevier and Wiley say that the site infringes their copyright, and ask the court to instruct Internet service providers in India to block access to it.

Sci-Hub's founder Alexandra Elbakyan argues that, in India, copyright is "not applicable in cases such as Sci-Hub, when [material] is required for science and education".



### Publishers unite to tackle doctored images in research papers

Legal experts say that there is a chance the court will rule in Sci-Hub's favour, because of a key aspect of the country's copyright law. The case hinges on the definition of 'fair dealings', which in the past has enabled institutions in India to lawfully reproduce academic textbooks and other copyrighted material for use in education.

If Sci-Hub wins, it could force publishers to rethink their business models in a similar way to how the music industry changed in response to the arrival of the Internet, says Arul George Scaria, a legal scholar at the National Law University, Delhi. Attitudes towards Sci-Hub in other countries could change on the basis of India's ruling, and the outcome could even influence similar cases in future.

## Pirate site

Previously, publishers have sued Sci-Hub and Elbakyan in several countries, and access has been blocked, or is due to be blocked, in 11 countries, including Germany, France, Sweden and the United Kingdom.

In lawsuits filed in recent years by Elsevier and the American Chemical Society, US judges ruled that Sci-Hub infringed the publishers' copyrights and owed them [US\\$15 million](#) and [\\$4.8 million](#), respectively. Elbakyan did not appear in court, or offer any legal representation for the site during those cases, and the fines have so far not been paid.

“Pirate sites like Sci-Hub threaten the integrity of the scientific record, and the safety of university and personal data,” the publishers behind the case in India told *Nature* in a statement. “They compromise the security of libraries and higher-education institutions, to gain unauthorized access to scientific databases and other proprietary intellectual property, and illegally harvest journal articles and e-books.” The publishers also allege that Sci-Hub uses stolen user credentials and phishing attacks to extract copyrighted journal articles illegally.

Elbakyan says that these are “empty accusations” that “have absolutely no content of evidence behind them”. She denies that Sci-Hub is a threat to science, or to the security of academic institutions. “Open communication is a fundamental property of science and it makes scientific progress possible. Paywalled access prevents this,” Elbakyan adds. “That is a threat, and not Sci-Hub.”

The site has proved popular among researchers, who say their institutions cannot afford costly journal subscriptions. India accounts for the third-

largest proportion of Sci-Hub's users, and when publishers brought the Delhi case in December 2020, a group of lawyers offered Elbakyan legal representation.

"There are serious questions of access to knowledge that the court ought to take into account," says Lawrence Liang, a legal scholar at Ambedkar University Delhi, who isn't part of the defence team but helped to rally support for Sci-Hub from scientists.

## Fair dealings?

The defence will argue that Sci-Hub's activities are covered by the list of exemptions in India's Copyright Act of 1957. One of these is that 'fair dealings' of a work can be used for private or personal use, including research.

Academic publishers have fallen foul of this section of the act before. In 2012, five publishers — including Oxford University Press and Cambridge University Press — unsuccessfully sued the University of Delhi and its photocopying shop for alleged copyright infringement in course packs made at the institution. These packs contained photocopies of passages and chapters from textbooks and, in some cases, copies of entire books that were produced for students, many of who could not afford to buy the originals.



## A guide to Plan S: the open-access initiative shaking up science publishing

The judge ruled that the university and the photocopying shop were not infringing the copyright of the books' publishers, because one of the exemptions listed in the copyright act includes reproducing work "by a teacher or pupil in the course of instruction". A key part of the case was evidence submitted to the court by students and teachers stating the need for the photocopies. This was allowed because there was deemed to be sufficient national interest in the ruling.

Liang was involved in that case, and says that India's fair-dealing provisions could be broad enough to facilitate the kind of access that Sci-Hub gives to articles. As with the textbooks, national interest in the case means that affected parties can submit evidence to the court. Earlier this year, 20 of India's top scientists argued that the country's scientific community "stands to be gravely prejudiced" if the case goes against Sci-Hub.

The scientists say in a document — known as a petition — submitted to the court that the case could have an "adverse impact on access to scientific knowledge, and so on science and technology research in India".

"Access to information is crucial for researchers. When the information is hidden behind paywalls, that curbs innovation," says Shahid Jameel, a virologist currently at the University of Oxford, UK, who signed the petition. Computational biologist Rahul Siddharthan at the Institute of Mathematical Sciences in Chennai, India, adds that "apart from a small number of elite institutes in India, most cannot afford to subscribe" to journals.

Further petitions supporting Sci-Hub have been submitted by medical doctors and policy advisers who use scientific papers as part of their work.

## **Ripple effect**

The case's next hearing is scheduled for 16 December, but legal experts warn that it could rumble on for years. Scaria says that the outcome will depend on whose rights the judge focuses on under the copyright rules. "If the judge views the matter from the perspective of user rights under

copyright law, there is a high chance that Sci-Hub will win the case,” he says. But if the judge views the matter from the perspective of the copyright holder, the verdict might go against the site.

The ramifications for publishers if Sci-Hub wins are hard to predict, according to Sci-Hub’s lawyers Shrutanjaya Bhardwaj and Sriya Sridhar. “Courts in progressive nations frequently borrow principles from foreign jurisdictions, and it is possible that Sci-Hub’s victory before the Delhi high court will cause a global ripple effect,” they say. On the other hand, a loss for Sci-Hub could see many researchers and institutions that cannot afford journal subscriptions being “excluded from access to scholarly work”.

Elbakyan says that the case could change everything for Sci-Hub. Winning could bring opportunities to improve the site and extend its reach.

“Today, the perception of Sci-Hub [is that] it is an illegal project, and that is even not disputable, but a fact,” she tells *Nature*. “Victory will show the ‘fact’ to be merely an opinion.”

*Nature* **600**, 370-371 (2021)

doi: <https://doi.org/10.1038/d41586-021-03659-0>

## Updates & Corrections

- **Clarification 14 December 2021:** This article has been updated to clarify that comments from Sci-Hub’s legal team come from both Shrutanjaya Bhardwaj and Sriya Sridhar.

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03659-0>

- NEWS
- 10 December 2021

# DeepMind AI tackles one of chemistry's most valuable techniques

Machine-learning algorithm predicts material properties using electron density.

- [Davide Castelvecchi](#)



The AI predicts the distribution of electrons within a molecule (illustration) and uses it to calculate physical properties. Credit: DeepMind

A team led by scientists at the London-based artificial-intelligence company DeepMind has developed a machine-learning model that suggests a

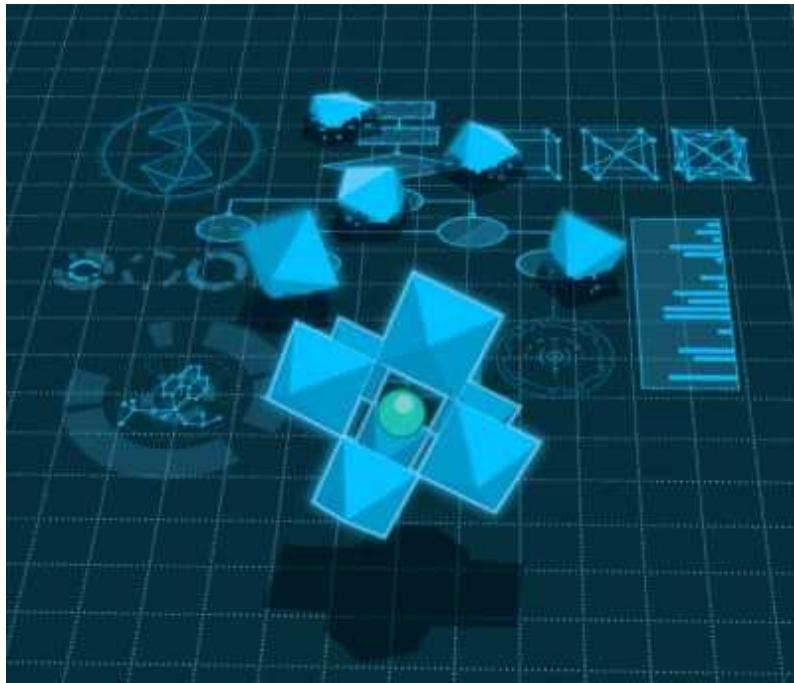
molecule's characteristics by predicting the distribution of electrons within it. The approach, described in the 10 December issue of *Science*<sup>1</sup>, can calculate the properties of some molecules more accurately than existing techniques.

“To make it as accurate as they have done is a feat,” says Anatole von Lilienfeld, a materials scientist at the University of Vienna.

The paper is “a solid piece of work”, says Katarzyna Pernal, a computational chemist at Lodz University of Technology in Poland. But she adds that the machine-learning model has a long way to go before it can be useful for computational chemists.

## Predicting properties

In principle, the structure of materials and molecules is entirely determined by quantum mechanics, and specifically by the Schrödinger equation, which governs the behaviour of electron wavefunctions. These are the mathematical gadgets that describe the probability of finding a particular electron at a particular position in space. But because all the electrons interact with one another, calculating the structure or molecular orbitals from such first principles is a computational nightmare, and can be done only for the simplest molecules, such as benzene, says James Kirkpatrick, a physicist at DeepMind.



## Can artificial intelligence create the next wonder material?

To get around this problem, researchers — from pharmacologists to battery engineers — whose work relies on discovering or developing new molecules have for decades relied on a set of techniques called density functional theory (DFT) to predict molecules' physical properties. The theory does not attempt to model individual electrons, but instead aims to calculate the overall distribution of the electrons' negative electric charge across the molecule. "DFT looks at the average charge density, so it doesn't know what individual electrons are," says Kirkpatrick. Most properties of matter can then be easily calculated from that density.

Since its beginnings in the 1960s, DFT has become one of the most widely used techniques in the physical sciences: an investigation by *Nature*'s news team in 2014 found that, [of the top 100 most-cited papers](#), 12 were about DFT. Modern databases of materials' properties, such as [the Materials Project](#), consist to a large extent of DFT calculations.

But the approach has limitations, and is known to give the wrong results for certain types of molecule, even some as simple as sodium chloride. And although DFT calculations are vastly more efficient than those that start from basic quantum theory, they are still cumbersome and often require

supercomputers. So, in the past decade, theoretical chemists have increasingly started to experiment with machine learning, in particular to study properties such as materials' chemical reactivity or their ability to conduct heat.

## Ideal problem

The DeepMind team has made probably the most ambitious attempt yet to deploy AI to calculate electron density, the end result of DFT calculations. “It’s sort of the ideal problem for machine learning: you know the answer, but not the formula you want to apply,” says Aron Cohen, a theoretical chemist who has long worked on DFT and who is now at DeepMind.

### DeepMind’s AI helps untangle the mathematics of knots

The team trained an artificial neural network on data from 1,161 accurate solutions derived from the Schrödinger equations. To improve accuracy, they also hard-wired some of the known laws of physics into the network. They then tested the trained system on a set of molecules that are often used as a benchmark for DFT, and the results were impressive, says von Lilienfeld. “This is the best the community has managed to come up with, and they beat it by a margin,” he says.

One advantage of machine learning, von Lilienfeld adds, is that although it takes a massive amount of computing power to train the models, that process needs to be done only once. Individual predictions can then be done on a regular laptop, vastly reducing their cost and carbon footprint, compared with having to perform the calculations from scratch every time.

Kirkpatrick and Cohen say that DeepMind is releasing their trained system for anyone to use. For now, the model applies mostly to molecules and not to the crystal structures of materials, but future versions could work for materials, too, the authors say.

*Nature* **600**, 371 (2021)

doi: <https://doi.org/10.1038/d41586-021-03697-8>

# References

1. 1.

Kirkpatrick, J. *et al.* *Science* **374**, 1385–1389 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03697-8>

| [Section menu](#) | [Main menu](#) |

- NEWS Q&A
- 03 December 2021

# Science misinformation alarms Francis Collins as he leaves top NIH job

The genome project leader reflects on his 12 years at the helm of a juggernaut biomedical agency, and what lies ahead.

- [Nidhi Subbaraman](#)



Francis Collins has been head of the US National Institutes of Health for longer than any other presidential appointee. Credit: Andrew Harnik/AP

Photo/Bloomberg/Getty

This month, Francis Collins [will step down as director](#) of the US National Institutes of Health (NIH) after more than 12 years leading the agency, the world's biggest public funder of biomedical research. A former head of the Human Genome Project, he championed similar bold, big-budget science efforts, such as the All of Us Project, which aims to study health data from one million people. He led the NIH under three US presidents, steered it through a roiling pandemic, and faced myriad clashes over politics and biomedical science. Collins will stay on at the NIH to continue research in his lab. He spoke to *Nature* about some highlights of his time at the helm, and issues facing the agency in the future.

## Which achievement will you cherish most?

It's really hard to pick one. Maybe it's where it was possible to bring together scientists of multiple disciplines and organize a truly bold, audacious project that would simply not have happened if one just counted on it coming together passively. I'm thinking of the [BRAIN Initiative](#). I'm thinking of what we did with [precision medicine and the All of Us Project](#), with the organized effort to develop COVID-19 vaccines in less than a year, and to develop diagnostic platforms for COVID-19 on similarly breathtaking timetables. Those are all things I feel proud of having played a role in.

## Are there projects you're disappointed you didn't get to finish?

Our efforts in diversity. I think we made some real progress there. I appointed my chief officer for scientific workforce diversity — and there's everything we're doing as far as recruiting, and as far as making sure our clinical trials are focused on diversity. But here we are in 2021, and if you look at our workforce, we are still woefully under-represented for people of colour. That's not what the NIH should look like. We are losing out on talent; we're losing out on productivity that we know comes from diversity. We've pushed hard on this during my time as NIH director, and we have made some progress, but we've got a long way to go.



Collins stands with former US president Barack Obama in 2013 as he announces the BRAIN Initiative, an enormous effort to map the brain and learn more about neurodegenerative diseases. Credit: Alex Wong/Getty

**US President Joe Biden's administration aims to launch ARPA-H, a multibillion-dollar agency, to accelerate research in science and health. It is supposed to be housed within the NIH. How would you gauge its success in two years' time?**

By then, we should certainly have recruited and installed a visionary director who has the appropriate attitude towards risk taking. That person should have been able to bring on board 50 to maybe 100 project managers who know how to identify projects that fit the ARPA-H model — projects that simply aren't getting done otherwise, and could be accelerated considerably if funded through this mechanism. I want to see some significant evidence of successes, but also some significant evidence of failures. Because if there

aren't some failures — and they need to fail early — then they're not being risky enough.

## **Observers have commended your ability to gain bipartisan support for the NIH among leaders in the US Congress. How do you do it?**

First of all, I've tried my darndest to stay out of any kind of political wrangling. I'm not a member of any political party. I've really endeavoured to make friends across the aisle and in both houses of Congress.

Relationships really matter. I have this incredibly positive message to share with members of Congress. I can come to a meeting — and I would bet I've had 1,000 of those in these 12 years — with information about medical research that may advance the cause of preventing or treating a terrible disease. And they're all concerned about that, for themselves, their families, their constituents. So most of these are incredibly positive experiences because of the topic, and because I'd make every effort to make the information accessible and not get all tangled up in a lot of jargon and complicated words.

## **What would it take for the next NIH director to succeed at this?**

It takes investment. The next NIH director needs to count on spending maybe a day every week interacting with leaders in the administration, and especially Congress — because they ultimately decide the budget — to build those relationships of trust. Then you know, when you sit down with a senator, that you've got this background of having had conversations — not always easy ones, because sometimes senators want things that we can't provide. But you also know that you're going to speak the truth to each other, and then everybody's bringing their best to that interaction and not trying to put one over on somebody.

## **The NIH faced criticism when it placed restrictions on fetal-tissue research during Donald Trump's presidency. Do you stand by the agency's actions?**

I think it's widely known that the NIH tried to protect the continued use of human fetal tissue. But ultimately, the White House decided otherwise. And we had no choice but to stand down. That is how the government works. Now that we have a different administration, that has changed.

One can learn from fetal tissue, important things that may ultimately save lives. There is the reality that in this country, pregnancy terminations are legal, and so there are lots and lots of fetal tissues being discarded every day. I've tried to make the case, as a Christian and somebody who really does think that human life is sacred, that it is more ethical to utilize some of these fetal tissues occasionally, in a way that might benefit somebody, than to put them into the incinerator. That doesn't win me friends in some constituencies, and it certainly didn't win the day in the Trump administration.

## **In 2018, He Jiankui shocked the world when he announced that twins had been born from CRISPR-edited embryos. Some researchers were aware of the work before it grabbed headlines. Do you think researchers have an obligation to raise an alarm about work that crosses legal or ethical lines?**

I do think they have a role. We are not just technicians. We are also supposed to be people who have a moral compass, and if something is happening in biological research that crosses that line into territory, which, in general, we as the human species have concluded shouldn't be happening, then it's up to us to point that out. I was not one of those who knew that this

experiment was going on. The whole thing was complicated by the fact that we don't have an international body that basically sets these kinds of ethical rules. This is one of the things that vexes me a bit. And I don't see it as likely that that's going to emerge anytime soon, given the fact that countries don't seem to be able to agree on a lot of things.



A musician, Collins often performed for staff, and at benefits and other venues during his time as NIH director. Here, he plays at the 13th Annual Prostate Cancer Foundation's Gala in 2017.Credit: Sean Zanni/Patrick McMullan/Getty

**Some say the US government's efforts to prevent espionage under the 'China Initiative' amount to racial profiling. The NIH has cooperated with these efforts and clarified guidelines for scientists reporting foreign funds and appointments. Should anything change in the government's approach?**

We're simply trying to identify places where people are doing things that are wrong. And recognizing that we have to act when we see that. I think we have to, when we approach what appears to be a troubling situation, try to start with giving the [research] investigator the benefit of the doubt that there was a careless failure to report something that should have been reported. I think that's what we are doing, and only when it becomes clear that there is an intentional distortion of the facts and intentional effort to physically hide information should we then be taking hard actions. We're working with our grantee institutions, because it's really their job to figure out what to do with their employees. We depend on them to decide what the appropriate action is.

**It has been reported that NIH grant recipient EcoHealth Alliance, based in New York City, might have conducted what some virologists would consider '[gain of function](#)' research conferring new abilities to coronaviruses. [Experts have said](#) the NIH allowed EcoHealth unusual latitude in that work. How would you characterize it?**

This term 'gain of function' has caused so much confusion and so much misunderstanding, some of it rather intentional, to distort the facts of what happened. Which is why we're trying to just avoid the use of that term. Let's talk instead about 'enhanced pathogens of pandemic potential'. [EcoHealth was] not crossing the line into the area that required that extensive kind of oversight, as was ultimately put forward in the US government's P3CO [[Potential Pandemic Pathogen Care and Oversight](#)] [guidelines](#). I think we, at this point, are incredibly transparent. It took a little while to get there because of concerns about setting precedents for revealing information that we normally don't share about interactions between us and the grantee. I will be quick to say [the research] has nothing to do with the origin of SARS-CoV-2. I don't know what more we could be sharing.

## You've watched science and politics collide for years. Do you believe politicization of science has grown worse?

It is much worse. And it's a reflection of the fact that polarization is much worse — and tribalism is much worse. We're in a really bad place. If science happens to produce a result that a political perspective doesn't like, then science has to be attacked. That's exactly what we see now happening, to the detriment of getting the facts out there.

## What role does the NIH have in pushing back against misinformation about science?

This has turned out to be a much more severe situation than I would have imagined a year ago. I wish we had more insights from behavioural social science research into how this has come to pass, and why it could have gotten so completely widespread. I want to call this out as one of my most major concerns as I stepped down from the NIH, of looking at the situation in our nation. Somewhere along the way, our political hyperpolarization began having a lot of really dangerous consequences, where in many instances we seem to have lost a sense of how to tell the difference between a fact and an opinion — or some Facebook post that's, frankly, a lie. That's truly dangerous. That's another epidemic that is not going to go away even if we triumph over COVID-19. We need to figure out what happened here, and how to bring ourselves back to a place where our nation has a more stable future.

*Nature* **600**, 372-373 (2021)

doi: <https://doi.org/10.1038/d41586-021-03611-2>

This interview has been edited for length and clarity.

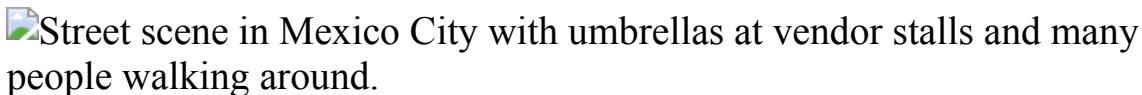
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 13 December 2021

# How the mixed-race mestizo myth warped science in Latin America

Researchers are trying to dismantle the flawed concept of homogeneous racial mixing that has fostered discrimination in Mexico, Brazil and other countries.

- [Emiliano Rodríguez Mega](#) <sup>0</sup>



Genetic studies have found a striking amount of diversity among people in Mexico. Credit: Stephania Corpi Arnaud for *Nature*

Nicéa Quintino Amauro always knew who she was.

She was born in Campinas, the last city in Brazil to prohibit slavery in 1888. She grew up in a Black neighbourhood, with a Black family. And a lot of her childhood was spent in endless meetings organized by the Unified Black Movement, the most notable Black civil-rights organization in Brazil, which her parents helped to found to fight against centuries-old racism in the country. She knew she was Black.

But in the late 1980s, when Amauro was around 13 years old, she was told at school that Brazilians were not Black. They were not white, either. Nor any other race. They were considered to be *mestiços*, or *pardos*, terms rooted in colonial caste distinctions that signify a tapestry of European, African and Indigenous backgrounds. And as one single mixed people, they were all equal to each other.

The idea felt odd. Wrong, even. “To me, it seemed quite strange,” says Amauro, now a chemist at the Federal University of Uberlandia in Minas Gerais and a member of the Brazilian Association of Black Researchers. “How can everyone be equal if racism exists? It doesn’t make sense.”

Amauro’s concerns echo across Latin America, where generations of people have been taught that they are the result of a long history of mixture between different ancestors who all came, or were forced, to live in the region.



Nicéa Quintino Amauro is a chemist in Minas Gerais, Brazil. Credit: Pablo Albarenga for *Nature*

In Mexico, for example, most people also think of themselves as mestizos, a term that emerged during the colonial period to explain the blend of ethnicities — especially between Indigenous peoples and the Spanish colonizers. The process of fusion, referred to as *mestizaje*, has been so intense that a lot of Mexicans say it no longer makes sense to talk about race or racism. The idea of a post-racial society drew important support from early genetic research in the twentieth century and modern human genomic studies, which show that most humans come from a mix of different

ancestries. This unifying vision took hold across Latin America, shaping public policies and conceptions about race.

But like all other race-based labels, the mestizo is a social construct, not a well-defined scientific category of people who share similar genetic characteristics. And many researchers have started to challenge the mestizo ideology, which they see as a source of pain for many people — and an obfuscating, sometimes troubling, influence in science.

“The mestizo is everything and nothing. It’s not very descriptive,” says Ernesto Schwartz-Marín, a Mexican anthropologist who probes the relations between genomics research and ideas of race at the University of Exeter, UK. “Why do we look at these categories, born out of the arbitrary nature of the colonial conquest, as biologically significant?”

The mestizo narrative suggests the absence of racism — even when there is ample evidence that skin colour is a powerful determinant of wealth and education levels across Latin America.

And the concept of *mestizaje* in effect suppressed the visibility and recognition of Indigenous and Black people in the region, while sometimes elevating European ancestry. The concept is another harmful vestige of colonial rule, says Jumko Ogata, who is pursuing a degree in Latin American studies at the National Autonomous University of Mexico (UNAM) in Mexico City. She is also an anti-racism educator and identifies as an Afro-Japanese Mexican. “It has romanticized and misrepresented many histories of dispossession, extraction, violence and colonization,” she says. “It was, and it still is, a profoundly violent thing.”

Her thoughts are part of a broader discussion that’s going on in Latin American society — and in the research community. Critics say that the mestizo myth has had a troubling influence in science. The way the mestizo label has been used in many genetic studies has misrepresented or ignored the histories of people with Indigenous or African ancestries, according to some researchers and activists. What’s more, they argue that this and other categories used in human genetics carry outdated or even racist perspectives.



Jumko Ogata is a student at UNAM in Mexico. Credit: Paulina Fi Garduño

Some call for banishing the mestizo label in human genetics and adopting much more specific terms not connected with colonial concepts. Others say that the mestizo idea is not as problematic as critics argue.

One thing is clear, says Vivette García Deister, an ethnographer of science at UNAM, who has studied how mestizo ideology has influenced genetic studies in Mexico. “There’s no easy solution.”

## Birth of an idea

The idea of *mestizaje* has come full circle from its origins. In the early 1900s, politicians across Latin America began to realize that this concept could be an effective tool for forging a national identity. At the time, ‘scientifically’ backed racism was prominent in the United States, where pseudoscientific eugenics theory was used to justify stripping people of their rights; this included the forced sterilization of thousands of African American, Native American and Puerto Rican people — which continued into the 1970s. Proponents of eugenics believed that mixing of races could result in ‘degeneration’ and ‘decay’. And they pointed to events south of the border to support their ideas.

The political turmoil that plagued post-independence Latin America was often attributed to the widespread mixing that had occurred since colonial times, and US politicians saw it as an opportunity to intervene, says Juliet Hooker, a Nicaraguan political scientist at Brown University in Providence, Rhode Island. “The idea was that, due to their racial heterogeneity, Latin American peoples were prone to chaos,” she adds. “They couldn’t really govern themselves, so they needed US intervention.”

But some Latin American intellectuals pushed back with the idea that racial mixing was in fact positive. And they formulated theories that encouraged Latin Americans to take up the mestizo identity, which they said combined the best features of each group. Their message was one of social cohesion, which many of these countries used to resist US domination and bring people together under a nationalistic sentiment.

“The mestizo myth was Latin America’s project to put an end to the idea that everything hybrid, everything mixed, was inferior,” says Schwartz-Marín. “It was born as a response to the racist purism of that time.”

This mixed-race ideology has prevailed since the 1920s and 1930s, although not without its harmful aspects. The blend of ethnicities was seen by privileged members of society as a way of ‘whitening’ the nation in the long run. And it disseminated the belief that the scourge of racism could not exist where the boundaries of human differences had been blurred so much.



Juliet Hooker is a Nicaraguan political scientist at Brown University in Rhode Island. Credit: Nick Dentamaro/Brown University Office of Communications

“Historically, these ideas serve to deny the presence of Indigenous peoples and Afro-descendants. To say that they no longer exist, that they have been absorbed by the process of *mestizaje*,” says Hooker, who experienced this as a girl when her family moved from the Afro-Caribbean coast of Nicaragua, where she grew up, to its mostly mestizo capital. The people there rarely identified as Black, even the ones who looked like her, and repeatedly asked

why she identified that way. In 2017, Hooker explored the origins and history of the mestizo myth in her book *Theorizing Race in the Americas*.

In the 1960s, the United Nations cultural organization UNESCO recognized that there is no genetic evidence for the existence of human races. Still, scientists were interested in studying race mixture, sometimes in ways that supported racist ideas at the time. And Latin America was seen as the perfect real-world laboratory in which to do so.

Brazilian geneticists were particularly productive in the field, and published a large number of papers that tried to explain the ‘trihybrid’ origin of Brazilians with Black, white and Indigenous ancestries; and researchers hunted for ‘racial markers’ to calculate degrees of race mixture.

In Mexico, physician-turned-geneticist Rubén Lisker used genetic markers to map enzyme deficiencies and abnormal haemoglobins, the body’s oxygen-carrying proteins, in what he called “the Indians, the descendants of the Spanish and their mix”. His approach followed the dominant ideology in Mexico, which emphasized mixture between Indigenous people and Europeans, and largely ignored contributions from people of African heritage — although his results did suggest that in certain parts of the country, between 5% and 50% of the genetic variation found in Indigenous populations was also found in African populations<sup>1</sup>. (During the colonial slave trade, millions of people from Africa were enslaved and taken to Latin America against their will.)

It took many years for Mexico to acknowledge the African connection. In 2015, after a long struggle for recognition, more than 1 million Afro-Mexicans were allowed to identify as such in the national census, which had previously not recognized them. By 2020, the number had increased to 2.5 million people, or 2% of the country’s population.

## Mestizo genomics

Until May 2007, Luiz Antônio Feliciano Marcondes never thought his identity would be dissected all across Brazil. But that month, media coverage exploded around his genetic ancestry. As one of the country’s

leading samba singers and composers, his Blackness had never been called into question. People needed only to see his stage name, Neguinho da Beija-Flor; ‘neguinho’ refers to his dark complexion and to Brazilians can mean an expression of affection, racism or both.

After examining 40 regions in Marcondes’s genome, a team of researchers led by human geneticist Sérgio Pena at the Federal University of Minas Gerais in Belo Horizonte, concluded that Marcondes had a predominantly European ancestry, at 67%. The result came as a surprise. “European, me?” he said in an interview. “I go by the colour of my skin.”



### Facing up to injustice in genome science

The ancestry test was part of a project commissioned by the broadcaster BBC Brazil to reveal the genetic profiles of nine Black celebrities, which showed many of them to have a mixture of different ancestries. These findings belong to a larger body of research that, since the 1990s, has strongly portrayed Brazil as a diverse nation that is also raceless, where everyone is so mixed that it is now impossible to differentiate people genetically into different groups or populations. But the idea of a raceless society conflicts with the lived experiences of many Brazilians, particularly those who don’t identify as white.

The thinking about racial mixing has progressed in different ways across Latin America. In Colombia, although the dominant view casts the country as one mixed nation, there is a strong tendency to divide it into several regions, each one with various types and degrees of admixture. In the Antioquia region, where Spanish colonizers established their first settlements, some people push an almost mythological narrative that exalts their European, white origins, says María Fernanda Olarte-Sierra, a Colombian ethnographer of science and technology at the University of Amsterdam in the Netherlands.

So when a paper in 2006 revealed that the ancestry of Antioquians is between 70% and 80% European<sup>2</sup>, the highest in the country, many people welcomed the findings that essentially reinforced the view of themselves as mostly non-Indigenous and non-Black, Olarte-Sierra says. But she notes that scientists only sampled people from the highlands, who mostly identify as mestizos, and left out Afro-Colombian communities living in Antioquia's coastal area and river valley. "I thought it was a self-fulfilling prophecy," says Olarte-Sierra. "It had consequences that went unquestioned."

Another prophecy was self-fulfilled in 2009 when Mexico's National Institute of Genomic Medicine (INMEGEN) revealed the results<sup>3</sup> of the government-led Mexican Genome Diversity Project. After a four-year effort to collect and analyse blood samples from people living across the country, scientists showed that most of the Mexican population derived from a mixture of Indigenous and European people, the two groups typically seen as contributing almost entirely to the mestizo majority. "This confirmed, on a molecular level, which were the ancestral populations of most Mexicans," says Gerardo Jiménez Sánchez, a medical geneticist who oversaw the project and was the founder and first director of INMEGEN.



Vivette García Deister is an ethnographer of science at UNAM. The mural says “Never again a science without us” in Spanish. Credit: Stephania Corpi Arnaud for *Nature*

The findings were a major event in Mexico, particularly because multi-ethnic populations had not been included in any of the global efforts to map genetic variation between people, such as the HapMap project, which aimed to capture common variations in the human genome. Jiménez Sánchez announced the results during a state ceremony at the presidential residence and presented them as ‘the book of life’ of Mexicans.

The event left some researchers feeling uncomfortable. “I found problematic that the mestizo ideology was not being questioned,” says García Deister. Although the project set out to measure diversity in Mexico, the study initially included a limited sample of 300 people labelled as mestizos and compared them with 30 people identified as Zapotecs, who represented Indigenous ancestry. Geneticists also sampled individuals from other Indigenous groups, but many were ultimately considered to have significant genetic admixture and they were excluded as ‘genomic noise’, says García Deister, who interviewed several of INMEGEN’s researchers. “Once again,

Indigenous populations were somewhat at the service of the nation,” she says, “because we needed them to tell us something about mestizos.”

INMEGEN’s project had a strong focus on genomic medicine, and Jiménez Sánchez says it provided huge amounts of data that have since helped geneticists to understand how Mexican mestizos respond to pharmaceutical drugs and which gene variants, many of them linked to Indigenous ancestry, are related to complex conditions such as diabetes or high blood pressure. “Studying that combination of genes with different ancestries can establish risks for diseases — or, better still, the possibility of protecting you from them,” says immunogeneticist Julio Granados Arriola at the Salvador Zubirán National Institute of Health Sciences and Nutrition in Mexico City.

But some worry about this attempt to find risk alleles and link them to certain populations of Indigenous ancestry. This essentially blames Indigenous people for health problems seen in Mexican mestizos, says Jocelyn Cheé Santiago, a Zapotec genomic scientist who is studying the philosophy of science at UNAM. “You stigmatize a whole population.”



Indigenous genomics researcher Jocelyn Ch  e Santiago in front of the library at the National Autonomous University of Mexico in Mexico City.Credit: Stephania Corpi Arnaud for *Nature*

Those studies might be seen as supporting the idea that disorders such as diabetes and obesity, which affect millions of Mexicans, are partly due to diets or lifestyle, but mostly to genes inherited from their Indigenous ancestors, says Peter Wade, a social anthropologist at the University of Manchester, UK, who has studied race issues and mestizo genomics across Latin America. “It’s kind of [implying] these health problems are somehow the fault of the Indigenous people or Indigenous ancestry.”

Jim  ez S  nchez did not respond to specific questions about these criticisms. “It is always very useful to know other points of view,” he said. “But those concerns that were genuine at the beginning have not occurred. They didn’t come true. I don’t see them.”

The medical focus on mestizo genetics might also create false hopes. At the beginning of the COVID-19 pandemic, Granados Arriola publicly affirmed that the mosaic-like genome of Mexicans would act as a barrier to slow the spread of the virus. The hypothesis, he explained in May 2020, relies on the presence of protective variants inherited by the Indigenous ancestors who survived the plagues brought by European settlers. He continued to support that hypothesis this year. “What I expect is that most Mexicans with Indigenous genes, regardless of other factors, will have greater protection” against severe disease, Granados Arriola told *Nature* in June.

So far, little evidence exists that specific genetic ancestries account for lower or higher rates of coronavirus infections<sup>4</sup>. Today, Mexico has one of the world’s highest COVID-related death tolls, with a total of more than 295,000 deaths.

Back in Brazil, Marcondes’s ancestry test indicating a strong European component figured in fierce public debates about who qualifies as Black. And previous genetic data produced by Pena and his colleagues took centre stage in the conversation.



### Don't ignore data from minority populations

Since the 1980s, as Pena began his research on paternity tests, he was struck by the “tremendous” genetic diversity he saw in the samples he examined. For more than two decades, Pena’s studies<sup>5</sup> have repeatedly shown that the genome of each Brazilian is a unique collage assembled from three ancestral groups: European, Indigenous and African populations. In colonial times, this was typically the outcome of European men raping women of Indigenous and African heritage.

Pena’s research has also demonstrated that admixture has slowly but surely uncoupled genetic ancestry from physical traits such as skin pigmentation, a finding that has been confirmed by other research groups<sup>6</sup>. This means that one cannot safely predict another person’s skin colour from their genetic ancestry, or vice versa.

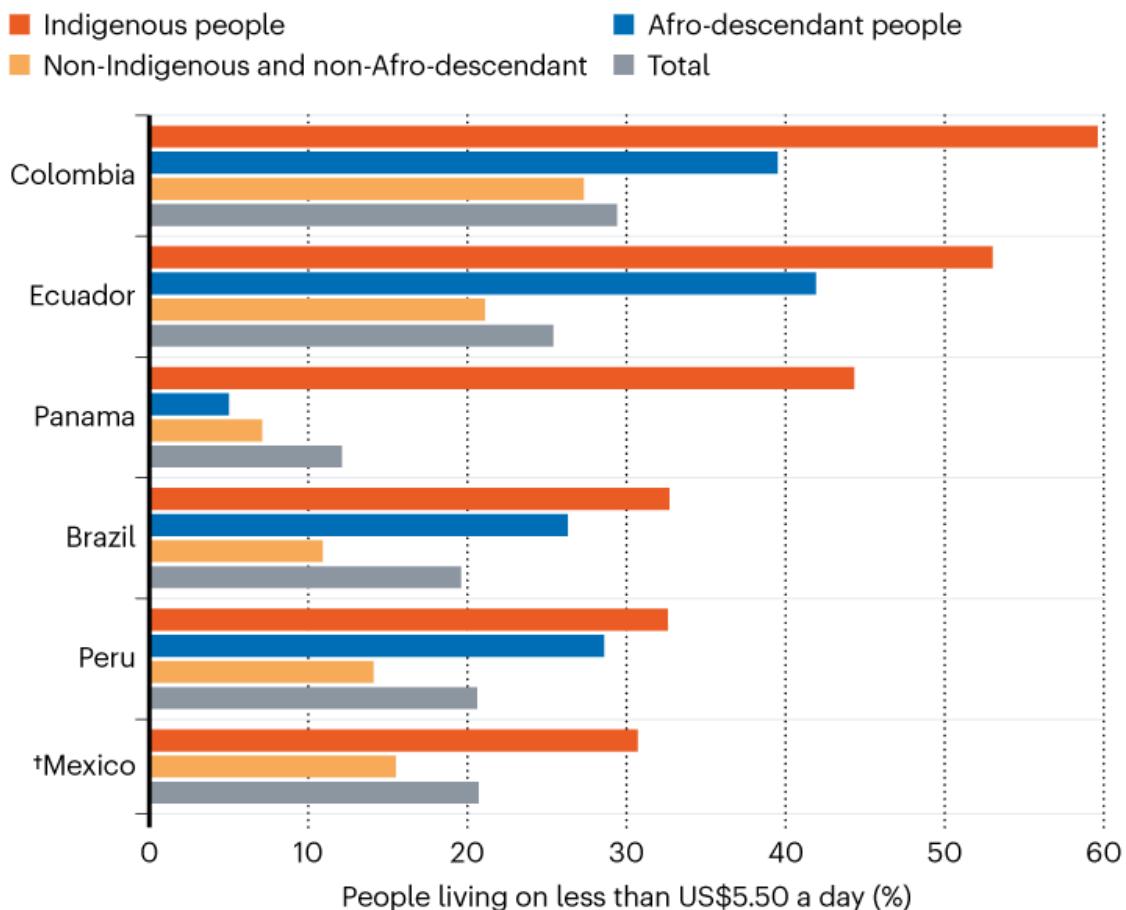
The evidence convinced Pena that “there are no human races and that racism is stupid”. He adds that his work has been guided by an anti-racist agenda. “My idea is that if Brazilians realize how much they owe to their Amerindian and African ancestry, we should be able to abolish racism in Brazil.”

Other people latched onto his results, particularly those supporting the biological non-existence of race and those showing generalized mixture

among Brazilians. In 2009, these studies were used by the centrist political party Democratas to support a petition to the country's supreme court asking it to revoke racial quotas for access to public universities. Pena's genetic data also inspired an anti-quota manifesto. The document, signed by 113 geneticists, social scientists, politicians, lawyers and citizens, argued that such policies represented a model imported from the United States, and were inappropriate for a country as mixed as Brazil. This was despite evidence that Afro-descendant people (that is, those with African heritage) generally face more inequalities in health care, education, employment, income and housing in Brazil and other Latin American countries (see 'Inequality in Latin America').

## INEQUALITY IN LATIN AMERICA

In many Latin American countries, rates of poverty are much higher for people of African or Indigenous heritage compared with other groups\*, a challenge to the idea of raceless societies.



\*Ethnicity based on self reported data; ethnic-group labels based on World Bank designations. †Data on Afro-descendants in Mexico are not available.

Source: World Bank

Pena himself signed the manifesto and participated in the supreme court hearings, giving evidence on the non-existence of race at the genetic level. But he defends his position as neutral. “Our genomics work on Brazilians is primarily descriptive and, as such, it is essential to understand who we are.”

In 2012, the court finally declared that genetics was irrelevant when it came to affirmative-action laws, but some researchers think the legacy of this event has been hurtful. “Our issues are not genetic,” protests Amauro, the Brazilian chemist. “To be discriminated against in Brazil, it’s enough for you to have a certain skin coloration, a certain hair texture, a certain jaw, nose, mouth.”

## A path forward

As genomic science in Latin America grappled with the legacy of mestizo ideology, some scholars decided to dissect that relationship and chart a path forward. “We were interested in how science was becoming involved with, challenging, reinforcing, changing ideas about the nation and ideas about race mixture,” says Wade.

In 2010, he teamed up with 13 collaborators in Brazil, Mexico and Colombia to launch a project that explored how concepts of race have influenced genomic research and vice versa, sometimes in problematic ways. The researchers spent more than two years interviewing and shadowing scientists in their laboratories, scrutinizing the methods and language they used and investigating how the results were disseminated. Their findings, compiled in a 2014 book called *Mestizo Genomics*, speak to continued concerns.

When scientists in the latter half of the twentieth century abandoned the racial categories used by their predecessors, they did so because the data showed that those artificial boundaries simply do not exist. Race, they concluded, has no biological foundation and does not come near to representing the convoluted history of our species. Instead, genetics used a wide range of genetic markers and statistical methods to more accurately reflect the differences and similarities archived in our genes, and found that,

in many cases, the resulting patterns broadly coincided with geography — giving rise to clusters that described African, European or Asian ancestries (the term Amerindian, also widely used, was originally created as a linguistic category).



### [Caltech confronted its racist past. Here's what happened](#)

But according to Wade and his team, these ancestry labels still evoke arbitrary, race-like ideas and categories that were invented in colonial times. The label of mestizo, mentioned countless times by geneticists in the papers they write, also inadvertently resurrects the concept of race that science has worked so hard to lay to rest, Wade says.

“If you’re talking about mestizos, you definitely are talking about race,” says Wade, adding that the mestizo, as a concept, has always symbolized a mixture between white colonizers, Indigenous communities and sometimes enslaved Africans and their descendants. “Every time you talk about mestizos, you’re automatically invoking the existence of those categories.”

Ethnographers and anthropologists also take issue with how geneticists have tended to use Indigenous or Afro-Latino populations, considered to be essentially homogeneous, as outside groups or reference points against which to compare mestizo samples. Sometimes this is done to understand how mixing occurred. But critics allege that the practice replicates the

nationalistic divide that has defined these groups as separate — and they say that genetic studies seem to incorrectly assume that genetic admixture has not occurred in non-mestizo populations.

“That drove me crazy,” says Schwartz-Marín. “Who is mestizo? Well, it depends on who you consider more or less mixed.”

Change, however, is not a quick process. “Sometimes it’s difficult to get rid of the term overnight,” says Andrés Moreno Estrada, a population geneticist at the National Laboratory of Genomics for Biodiversity in Irapuato, Mexico. In 2014, he and his colleagues set out to map the genetic blueprint of Mexicans, only to find a striking amount of diversity in the country — with some Indigenous groups in Mexico being as different from each other as Europeans are from East Asians<sup>7</sup>. Their results also suggested something else: mestizos and Indigenous groups are pretty much the same at the DNA level.

“If you take a mestizo and an Indigenous person from the same region, genetically they are indistinguishable,” Moreno Estrada says. “This dichotomy has no biological basis.”

Mestizo is not the only category that has gone relatively unexamined in human genetics. An analysis of biomedical papers since 2010 found that nearly 5,000 of them used Caucasian — an eighteenth-century term deeply rooted in racism and genetically meaningless — to describe certain populations.



### Too many scientists still say Caucasian

“The way that we name the [genetic] clusters that we see affects people’s understanding of race,” says Jennifer Raff, an anthropological geneticist at the University of Kansas in Lawrence. In June she co-authored a preprint<sup>8</sup> about how some widely used categories, such as Caucasian, carry scientifically misleading or even racist perspectives. The goal, says Raff, is to stimulate a much-needed conversation about the language of genetics. “Let’s really dig down — let’s really talk about what we’re doing and how we’re doing it,” she says, “and what that means.”

Although the category mestizo is still in use, some geneticists have chosen to replace it with different terms, such as ‘admixed population’, ‘population of mixed origin’ or ‘cosmopolitan population’. Moreno Estrada says this at least strips away the historical and political baggage and the social implications of *mestizaje*. But others have come up with entirely new taxonomies.

When Colombian geneticist and statistician William Usaquén travelled to the Colombian La Guajira desert, at the northernmost tip of South America, to describe the genetic composition of the people there, he soon realized that the standard, clear-cut categories of genetic ancestry would not be good enough. The region, he learnt, is the home of the Indigenous Wayúu people, but other communities who do not consider themselves Indigenous, such as

the *guajiros*, have also lived there for generations. What's more, since the 1800s, the area has been a smuggling route, which attracted migrants from other parts of Colombia who settled. Usaquén found that some people had multiple ancestries and others belonged to families that had always identified as either Wayúu or *guajiro*.

"By this point, the categories of Amerindian or mestizo didn't fit anywhere," says Usaquén, who works at the National University of Colombia in Bogotá. He and his team decided to try something new. They came up with a system of seven different categories that accurately reflected the genealogy, history and demography of the population, and were able to determine how each of them genetically interacted with each other<sup>2</sup>. "There's no standard category you can use," adds Usaquén, who found that the Wayúu are a highly admixed group, but still distinct from other people. When thinking about which category to use, he says, "the big conclusion for us is that every time you study a population, you have to create them".

That's much easier said than done. Even if geneticists come together to rethink the language they use, "Then what?" says Schwartz-Marín. "Every new category we generate will have its own implications."

In Latin America, the myth of *mestizaje* maintains a foothold in different aspects of modern society, including science. And some researchers think it could be time to abandon the idea — particularly those who do not see themselves as fitting into the mestizo narrative.

Amauro is one. She resents the message that all Brazilians constitute one diverse, yet uniform, mixture. "When I put myself in this group where everyone is the same, my characteristics are lost," she says. "I can be anyone else. And when I am anyone, I am not me."

*Nature* **600**, 374-378 (2021)

doi: <https://doi.org/10.1038/d41586-021-03622-z>

## References

1. 1.

Suárez-Díaz, E. *Stud. Hist. Philos. Biol. Biomed. Sci.* **47**, 108–117 (2014).

2. 2.

Bedoya, G. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 7234–7239 (2006).

3. 3.

Silva-Zolezzi, I. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 8611–8616 (2009).

4. 4.

Shelton, J. F. *et al.* *Nature Genet.* **53**, 801–808 (2021).

5. 5.

Pena, S. D. J., Santos, F. R. & Tarazona-Santos, E. *Am. J. Med. Genet. C* **184**, 928–938 (2020).

6. 6.

Kim, J., Edge, M. D., Goldberg, A. & Rosenberg, N. A. *Am. J. Phys. Anthropol.* **175**, 406–421 (2021).

7. 7.

Moreno Estrada, A. *et al.* *Science* **344**, 1280–1285 (2014).

8. 8.

Birney, E., Inouye, M., Raff, J., Rutherford, A. & Scally, A. Preprint at arXiv <https://arxiv.org/abs/2106.10041v1> (2021).

9. 9.

Rojas, M. Y., Alonso, L. A., Sarmiento, V. A., Eljach, L. Y. & Usaquén, W. *Ann. Hum. Biol.* **40**, 119–131 (2013).

---

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03622-z>

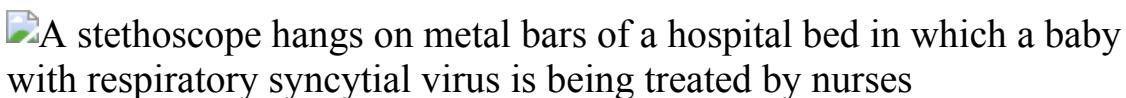
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 10 December 2021

# The race to make vaccines for a dangerous respiratory virus

Millions of people a year are hospitalized by respiratory syncytial virus and tens of thousands die. After decades of failure, four vaccines are now in late-stage trials.

- [Kendall Powell](#) <sup>0</sup>



An infant receives ventilation for respiratory syncytial virus. Infections are often worst in young babies. Credit: Marijan Murat/dpa/Alamy

In February 2020, just as a new coronavirus was triggering the global COVID-19 pandemic, structural biologist Jason McLellan and his team published the structure of the key protein it uses to invade human cells<sup>1</sup>. Immediately, scientists began [using that protein's structure to develop COVID-19 vaccines](#).

But that wasn't the first time McLellan, now at the University of Texas at Austin, had solved a structure and spurred on a new wave of vaccines. In 2013, he was focusing on a different killer — respiratory syncytial virus, or RSV<sup>2</sup>.

RSV causes a respiratory tract infection that affects 64 million people per year worldwide. It hospitalizes 3 million children under 5 years old and approximately 336,000 older adults annually (see 'Common scourge'). The

global health-care costs of RSV-associated infections in young children in 2017 were estimated to be US\$5.45 billion<sup>3</sup>.

Researchers have been trying for decades to develop a vaccine, and have had some particularly devastating failures — including the deaths of two participants in a trial in the 1960s.

Solving the protein structure revived the RSV field. McLellan, then a postdoctoral researcher at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, and his colleagues looked at a protein that the virus uses to fuse with cells and infect them, called the F protein, and found a way to stabilize it in its prefusion form — the shape it adopts when ready to grab on to cells. The structure of the prefusion F protein unveiled the best target for making vaccine-induced antibodies that could prevent the virus from entering human cells.

Now, an effective RSV vaccine is nearly within reach: four candidates and one monoclonal antibody treatment are in late-stage clinical trials.

“It’s been 8 years since the prefusion F protein conformation was elucidated and now we’re all in phase III trials based on that fundamental discovery,” says Christine Shaw, vice-president of early-development programmes in infectious diseases at messenger RNA therapeutics company Moderna in Cambridge, Massachusetts.

The quest for RSV vaccines has not escaped the effects of the COVID-19 pandemic. The pandemic has complicated the trials, but it also helped spur developments that might finally protect against this childhood killer.

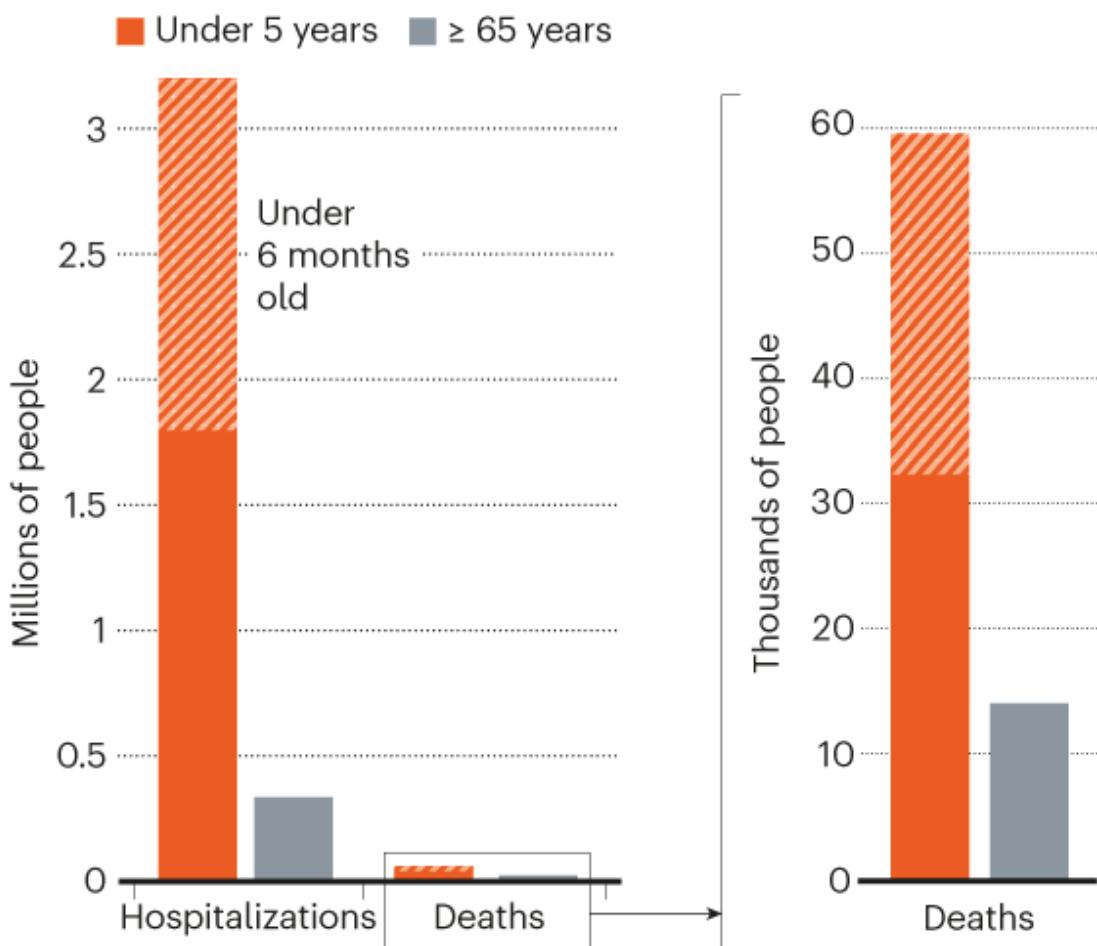
## Rough start

RSV infects most children by age three, and most adults many times over, but natural immunity to it is not long lasting. Infections are usually most severe in infants under two months old, who are encountering the virus for the first time. A vaccine or treatment would drastically reduce hospital and intensive-care admissions for this most vulnerable group, says Rabia Agha, a paediatric infectious-disease specialist at Maimonides Children’s Hospital in

New York City. And because of the acute medical need for vaccines and treatments for RSV, regulatory agencies in the United States and Europe are prioritizing them for review.

## COMMON SCOURGE

Despite most infections being mild, respiratory syncytial virus (RSV) hospitalizes millions of people a year\* and can be deadly, especially in the very young and older people.



\*Data shown are from 2015

©nature

Sources: Under 5, Shi, T. *et al. Lancet* **390**, 946–958 (2017); over 65, Shi, T. *et al. J. Infect. Dis.* **222**, S577–583 (2020)

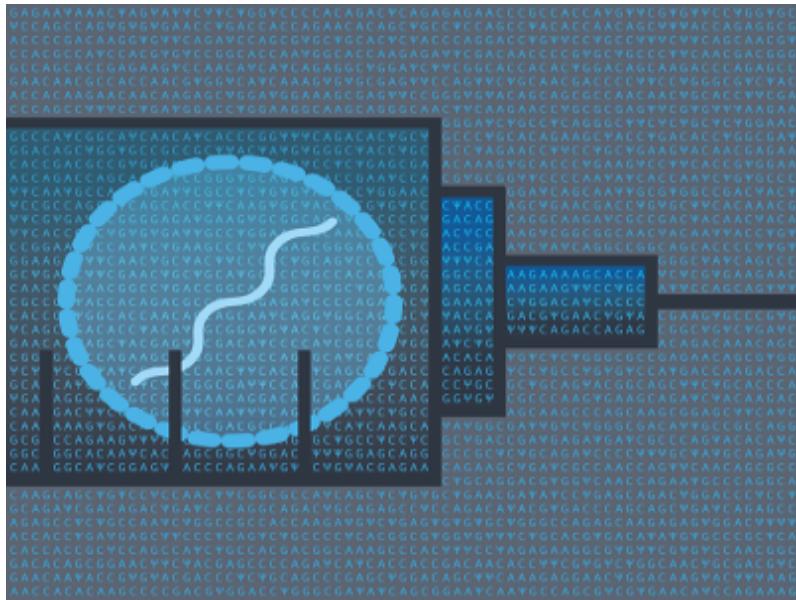
These candidates have been a long time coming. In the late 1960s, a series of vaccine trials tested a vaccine made from inactivated RSV in children in the

United States. Tragically, [the shot did the opposite of what it was meant to](#) — it made disease worse in vaccinated children when they were later naturally infected with RSV. This caused severe lung disease that hospitalized most of the children in the trial and killed two.

Steven Varga, a viral immunologist who studies RSV at the University of Iowa in Iowa City, says that early vaccine was “the culmination of a perfect storm of events”: it evoked too few virus-blocking antibodies and simultaneously caused an overactive inflammatory response. Researchers now know more about the science of both RSV infection and vaccinology, and Varga says there is no concern that the current crop of vaccine candidates could provoke a similar response.

Vaccine researchers soldiered on, but were targeting the wrong viral proteins — or the wrong forms of these proteins — for decades. “People have tried to make RSV vaccines since the 1970s, but it had been one failure after another,” says Rino Rappuoli, senior vice-president and chief scientist for vaccines at GlaxoSmithKline (GSK) Vaccines, based in Siena, Italy. “Things changed completely when we began to understand the main viral protein, protein F.”

In a very similar fashion to SARS-CoV-2 and its spike protein, RSV uses the prefusion form of the F protein (preF) to wedge its way into human cell membranes and fuse with them. Once that happens, the protein undergoes a major conformational shift to its more stable postfusion form (postF). Many of the earlier failed vaccines targeted postF.



## The tangled history of mRNA vaccines

By contrast, preF induces potent neutralizing antibodies that have been found to block the virus from entering human cells, Varga says.

To solve the structure of preF in a way that would make it useful for vaccine design, McLellan's team made a version of it in which the protein was fixed in a particular shape, so that its vulnerabilities were on display. To stabilize the protein in this shape, the team added a crucial chemical bond that acted like double-sided tape, keeping the protein folded in the preF shape, with its key antibody-target sites exposed. Vaccines that use this stabilized form prompt a much stronger antibody response from the immune system. All the current RSV vaccine candidates involve this stabilized form of preF, and healthy adults produce high levels of neutralizing antibodies against it.

Vaccines work by presenting the immune system with key bits of a pathogen, which induce the production of antibodies and immune cells that can recognize and fight the whole pathogen when the person is re-exposed to it. Many RSV vaccines use a similar array of technologies to those in development or use for COVID. Of the four RSV vaccines in phase III trials, two, from GSK and Pfizer, contain the stabilized preF protein itself. Another, from Janssen, uses a modified adenovirus that produces preF after delivery into the body, plus a dose of the pure protein, in the same shot. The

fourth, from Moderna, delivers modified mRNA that produces preF once the RNA is inside cells.

## Promising trials

Once approved, an RSV vaccine will be most useful for the very young and the very old. But to simplify clinical trials, all four companies in the race are initially testing their shots in people older than 60.

So far, early-stage data from phase II trials, some in older adults and some in younger adults, suggest that the vaccines will be safe and effective. Both Janssen and Pfizer have run small-scale challenge trials, exposing vaccinated younger adults to RSV, and have shown that their vaccines protect against infection. The Janssen, Moderna and GSK vaccines increased levels of neutralizing antibodies 9–15-fold.

GSK, Janssen, Moderna and Pfizer now each have global phase III trials under way in tens of thousands of older adults. GSK leads the pack, and Rappuoli says he expects to have interim results in early 2022.



### [How protein-based COVID vaccines could change the pandemic](#)

Newborn babies' immature immune systems present a more complicated challenge. They don't respond robustly to many vaccines, which is why

most childhood vaccinations are given after two months of age — just outside a critical window for the most severe RSV infections in infants. Instead, babies can be immunized while still in the womb. A few months before birth, a pregnant person is given the vaccine; their body makes antibodies and these are transferred through the placenta, and later through breast milk, to their newborn.

GSK and Pfizer are running phase III trials of their vaccines in pregnant people and following their babies to test for antibody levels at birth and for efficacy at preventing RSV infections during their first year.

In a good sign that this immunization strategy works, infants in Pfizer's phase II trial had higher antibody titres than their birth parent and the vaccine showed 85% efficacy at protecting babies against RSV infections requiring medical attention, says Alejandra Gurtman, vice-president of clinical research and development at Pfizer in Pearl River, New York.

A second way to protect newborns is by injecting them with antibodies that target the virus. AstraZeneca and Sanofi have partnered to test a monoclonal antibody called nirsevimab directed against stabilized preF, which has proved effective at reducing RSV infections in a phase III trial in healthy premature and full-term term infants.

Varga predicts that because older adults, pregnant people and the very young require different vaccine strategies, the vaccine race will not be winner takes all. "More than one of these RSV products will gain approval in each of these spaces," he says.

## Pandemic help and hindrance

Running clinical trials for an RSV vaccine during a pandemic caused by a different respiratory virus has brought both challenges and opportunities, vaccine developers say. One problem is that lockdowns and social distancing enacted to stop the spread of COVID-19 caused the RSV season — typically the winter — to shift considerably during 2020–21. Agha and her colleagues at the Maimonides Children's Hospital closely tracked cases of verified RSV, and found that they dropped to nearly zero in April 2020, staying that

low throughout the Northern Hemisphere winter and then peaking the following summer<sup>4</sup>. Such an unpredictable season makes it hard for health-care institutions to be properly prepared, and might mean it takes longer for participants in RSV vaccine trials to be exposed to the virus, which is needed to test the effectiveness of their jab. Whether the RSV season will return to normal remains to be seen, says Agha.

Nevertheless, the success and speed of the COVID-19 vaccines' development was a boon. "That success created awareness at every level, renewed interest to participate in clinical trials, and brought new platforms like mRNA vaccines that might be more efficient ways to stimulate the immune system," says Gurtman.

Moderna's mRNA-based RSV vaccine was in development before SARS-CoV-2 appeared, giving the technology a head start. "The COVID vaccine benefited from the RSV programme, and then it flipped," with the RSV programme benefiting from efficiencies gained from the COVID-19 programme, says Shaw.

Agha says the COVID-19 vaccines exemplified how quickly an effective vaccine can now be made against a respiratory virus. "As infectious disease experts, we're always excited when there is hope in the pipeline to make RSV a preventable disease in the future."

*Nature* **600**, 379–380 (2021)

doi: <https://doi.org/10.1038/d41586-021-03704-y>

## References

1. 1.

Wrapp, D. *et al. Science* **367**, 1260–1263 (2020).

2. 2.

McLellan, J. S. *et al. Science* **342**, 592–598 (2013).

3. 3.

Zhang, S. *et al.* *J. Infect. Dis.* **222**, S680–S687 (2020).

4. 4.

Agha, R. & Avner, J. R. *Pediatrics* **148**, e2021052089 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03704-y>.

| [Section menu](#) | [Main menu](#) |

## Books & Arts

- **[The loss of the world's frozen places](#)** [ 14 December 2021]

Book Review • Two very different books explore the past, present and future of glaciers.

---

- BOOK REVIEW
- 14 December 2021

# The loss of the world's frozen places

Two very different books explore the past, present and future of glaciers.

- [Alexandra Witze](#)



Argentina's glaciers are protected by a national law that restricts activities including mining in their vicinity. Credit: Alamy

**The Last Winter: The Scientists, Adventurers, Journeymen, and Mavericks Trying to Save the World** Porter Fox Little, Brown (2021)

**Meltdown! The Earth Without Glaciers** *Jorge Daniel Taillant* Oxford Univ. Press (2021)

In the summer of 1948, a small group of men in Juneau, Alaska, packed boots, axes, ropes and camping gear, and set off into the ice and snow that sprawls just north of the city. Led by glaciologist and US Navy veteran Maynard Miller, the explorers hiked through mountains and across crevasses, to see whether the frozen expanses preserved a long-term history of climate. They did. So Miller began an annual trip, taking teams onto the ice to chart the advance and retreat of glaciers, measure pollution falling on the snow and track other environmental changes.

Today, fresh cohorts of students — no longer just men — traverse 120 kilometres of the Juneau Icefield each summer. They travel using crampons and skis, to gather weather and climate observations. The two-month journey also forges future scientific leaders. Alumni include Steve Squyres, lead investigator for a pair of NASA's Mars rovers, who credits the experience with getting him into geoscience.



[Making the planetary personal: the roots of climate science](#)

Although interrupted in 2020 by COVID-19, the Juneau Icefield Research Program is one of the longest-running continuous studies of glaciers in North America. That puts it on the front lines of efforts to understand how

glaciers are responding to global warming — something Miller had only an inkling of seven decades ago.

The programme stars in writer Porter Fox's latest book, *The Last Winter*. The sorts of insights it has yielded about the cryosphere motivate *Meltdown!* by Jorge Daniel Taillant, who helped Argentina to adopt the world's first law to protect glaciers and permafrost. These books aim to add to the growing literature on the loss of the world's frozen places. In a crowded field, each succeeds only in part.

## Majesty and horror

Glaciers — flowing rivers of ice — are the largest source of fresh water on Earth, supplying billions of people in regions such as the Himalayas. Big polar glaciers, including Thwaites Glacier, are central to how rapidly the West Antarctic ice sheet is melting into the ocean, raising sea levels.

Glaciers collectively lost an average of 267 billion tonnes of ice each year between 2000 and 2019, and the figure is growing as the planet heats up.

The collapse of glaciers has been described visually: photographer James Balog's time-lapse videos, featured in the 2012 film *Chasing Ice*, reveal the majesty and horror of the loss. And several recent books tackle the topic, from explorer M Jackson's 2019 *The Secret Lives of Glaciers* — a study of the impacts of melting on a small Icelandic community — to journalist Dahr Jamail's paean *The End of Ice* (2019).



The Juneau Icefield in Alaska has been a training ground for Earth scientists for more than 70 years. Credit: Andrew Opila/Shutterstock

Fox moves into this space from his background as a ski writer (a previous book was about skiing and climate change). With *The Last Winter*, he aims to broaden the story of climate change beyond its impact on recreation. Yet he still focuses heavily on people's experience of winters in popular skiing destinations such as Washington state and the Alps. Fox asks questions about how the sorts of loss tracked by the Juneau programme will affect people living in developing nations — but provides no answers. Instead, his narrative hymns epic powder runs among Italian peaks, where breakfast at a mountain 'hut' includes Bavarian cream-filled doughnuts ferried in by snowmobile.

Fox does spend one of the four journeys that make up the narrative in Greenland, where he attempts to understand the impacts of climate change on Inuit peoples. Yet even here, his experience is mediated mainly by Western scientists, explorers and trip mates. His week on a dog-sledging

expedition with Inuit seal hunters is cut short when COVID-19 begins to rage outside Greenland.

Fox's narrative is one of eclectic adventures in Western experiences of the north; Taillant's is more textbook than travelogue.

A self-proclaimed cryoactivist, Taillant has spent decades working to protect the frozen environment. In the Andes in the early 2000s, he helped to lead a fight against a mining company that was dynamiting glaciers to access gold ore beneath. Today, he is an outspoken advocate of preserving glaciers through regional environmental-protection laws as well as international climate commitments.

## Hidden reserves

Taillant provides a thorough global overview of the extent of glaciers, the causes of glaciation and the many impacts of glacial disintegration, including dangerous floods. He does not interview scientists in the field, as Fox does in Juneau and elsewhere, but instead relies on news reports and other secondary material.



[180,000 forgotten photos reveal the future of Greenland's ice](#)

Particularly intriguing is a chapter on rock glaciers, which are common in the Andes and the American West but have been relatively little studied. These rubble-covered deposits, maybe 50% ice, typically exist at lower elevations than ice glaciers, in regions where surface sediments freeze and thaw but ice can remain year-round at greater depths. Rock glaciers can creep slowly downhill — in satellite images, they look like giant dirty flows of debris. Insulated by stony cover, they melt more slowly than ice glaciers, so could provide a slow, steady supply of water downstream.

Taillant also shines when he highlights innovative methods for conserving or even building glaciers. One scientist in Peru has covered glaciers with sawdust to slow their melt. In India, an engineer dubbed the Glacier Man has dammed snow in the winter so that it solidifies into artificial glaciers that release their water late into the year. Unfortunately, such messages come near the book's end, after much dense priming, unleavened by many exclamation marks and italics.

As we experience today's warming and prepare for even more, the future of glaciers and of winter is of huge concern. This year's United Nations climate-change conference, COP26, made clear that we are probably heading for well over 2 °C of warming. From the ice fields of Alaska, where future Earth scientists train, to the rock glaciers of the Andes that might hold precious, overlooked water resources, we need fewer empty promises — and more solutions.

*Nature* **600**, 381-382 (2021)

doi: <https://doi.org/10.1038/d41586-021-03705-x>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03705-x>

# Opinion

- **Decision makers need constantly updated evidence synthesis** [ 15 December 2021]  
Comment • Fund and use ‘living’ reviews of the latest data to steer research, practice and policy.
- **To beat Omicron, Delta and bird flu, Europe must pull together** [ 10 December 2021]  
Correspondence •
- **Chile: elect a president to strengthen climate action, not weaken it** [ 10 December 2021]  
Correspondence •
- **Prepare PhD holders for different career tracks** [ 14 December 2021]  
Correspondence •
- **Remembering a pioneer in biotechnology** [ 14 December 2021]  
Correspondence •

- COMMENT
- 15 December 2021

# Decision makers need constantly updated evidence synthesis

Fund and use ‘living’ reviews of the latest data to steer research, practice and policy.

- [Julian Elliott](#)<sup>0</sup>,
- [Rebecca Lawrence](#)<sup>1</sup>,
- [Jan C. Minx](#)<sup>2</sup>,
- [Olufemi T. Oladapo](#)<sup>3</sup>,
- [Philippe Ravaud](#)<sup>4</sup>,
- [Britta Tendal Jeppesen](#)<sup>5</sup>,
- [James Thomas](#)<sup>6</sup>,
- [Tari Turner](#)<sup>7</sup>,
- [Per Olav Vandvik](#)<sup>8</sup> &
- [Jeremy M. Grimshaw](#)<sup>9</sup>

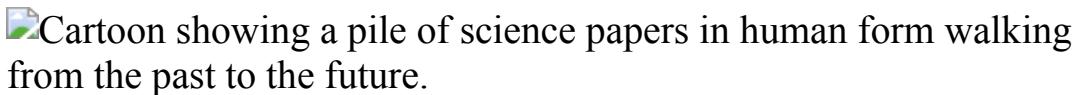


Illustration by David Parkins

As the world began to respond to the COVID-19 pandemic last year, there was an explosion of guidelines, position statements and protocols — many of low quality and contradictory. In March 2020, several of us approached the Australian National Health and Medical Research Council, worried that the cacophony would create confusion and anxiety among already-stressed clinicians. We argued for key bodies to come together quickly and use robust, evidence-based processes to find signals in the noisy flow of COVID-19 research. Two weeks later, we had formed a task force and

produced the first version of national, evidence-based COVID-19 guidelines for Australia. We made a commitment to update the guidelines every week, but this had never been done before. Our challenge was to work out how.

Typically, national guidelines draw on formal summaries of research evidence called systematic reviews, but the pandemic ‘broke the evidence pipeline’<sup>1</sup>. Take the example of remdesivir, an intravenous treatment originally developed for Ebola virus. In May 2020, weak but promising data suggested it could be used to treat COVID-19. Over the next 18 months, 52 papers from 14 randomized trials were published. Clinicians and policymakers had to make decisions on the basis of this shifting, and often contradictory, body of evidence. To help them, scholars produced systematic reviews — 30 in 2020. Many were out of date before publication because they left out recently published primary studies; most of the rest became out of date within weeks.

Out-of-date systematic reviews are common any time there is a flood of new research. In the absence of up-to-date summaries of accumulating knowledge, decision makers’ attention often jumps from study to study. This muddles policymaking, fuels controversy and erodes trust in science.

A better system would keep summaries of research evidence up to date. That’s the system we built for the [Australian National COVID-19 Clinical Evidence Taskforce](#), for the World Health Organization’s (WHO’s) [Therapeutics and COVID-19: living guideline](#), and for the [COVID-NMA repository](#) of COVID-19 research. The system can be applied beyond COVID-19, and indeed beyond health. Here’s what that will take.

## Evidence building

The process of identifying and combining data across studies to create a clear understanding of a body of research is known as evidence synthesis. The resulting publications are generally called systematic reviews, and often include a meta-analysis. Since the 1980s, the practice of evidence synthesis has grown to become the foundation for high-impact decision-making in disease prevention, diagnosis and treatment, and in other aspects of health. Evidence synthesis also helps to tackle questions in education, economics,

environment, criminal justice, global development and more. The number of published academic systematic reviews has risen from around 6,000 in 2011 to more than 45,000 in 2021.

University departments, international bodies such as Cochrane (a leading producer of systematic reviews on health topics), and numerous conferences and journals have established scientific methods, conventions and production systems for evidence synthesis. The approach is now routine across many national and global decision-making bodies responsible for policy and practice guidance. For example, rather than relying on single studies, drug regulators and insurers worldwide use syntheses of all relevant studies to evaluate safety and effectiveness and decide whether to approve or pay for a drug.

Evidence synthesis can be applied to the most pressing global challenges: climate change, energy transitions, biodiversity loss, antimicrobial resistance, poverty eradication and so on. But in practice, synthesis projects are often under-resourced, of poor quality, uncoordinated, duplicative and out of date.

‘Living evidence’ is an approach to synthesis that we and others have developed to address these problems<sup>2</sup>. It produces ‘ready-to-go’ evidence summaries that better serve the needs of decision makers because they are both rigorous (all relevant research has been carefully appraised) and up to date. Living evidence is particularly important in topics for which research evidence is emerging rapidly, current evidence is uncertain and new research might change policy or practice. For example, the Australian national guidelines for treating people who have had a stroke are now updated every 3–4 months instead of every 5–7 years (see ‘Evidence accelerated’). Clinicians can consult the most up-to-date evidence, and this has led to greater trust and use of the guidelines.

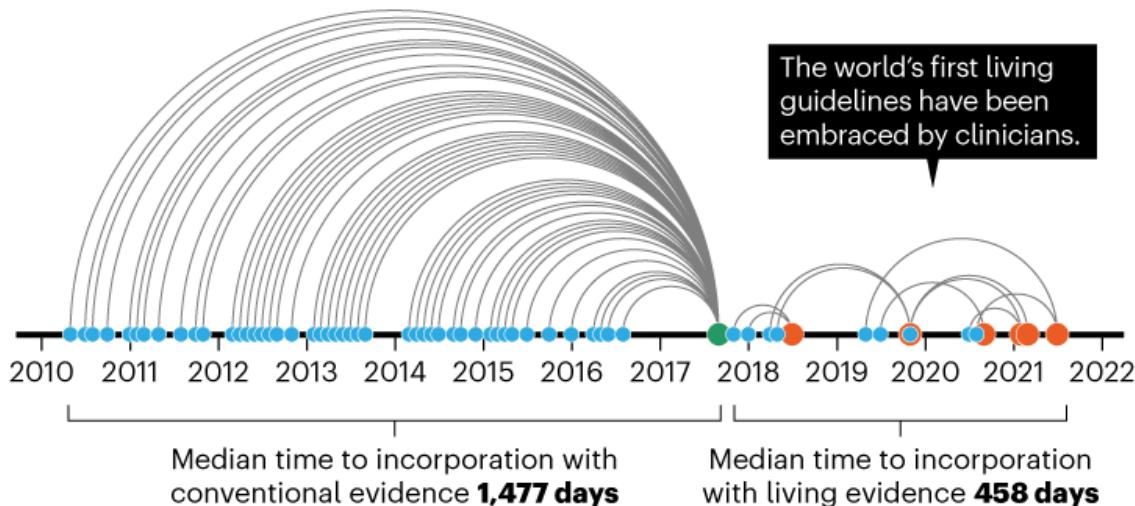
# EVIDENCE ACCELERATED

Using a living-evidence approach, researchers find, appraise and incorporate research in frequent cycles, rather than always starting from scratch.

- Primary study
- Guideline publication (conventional)
- Guideline publication (living)
- Time to publication

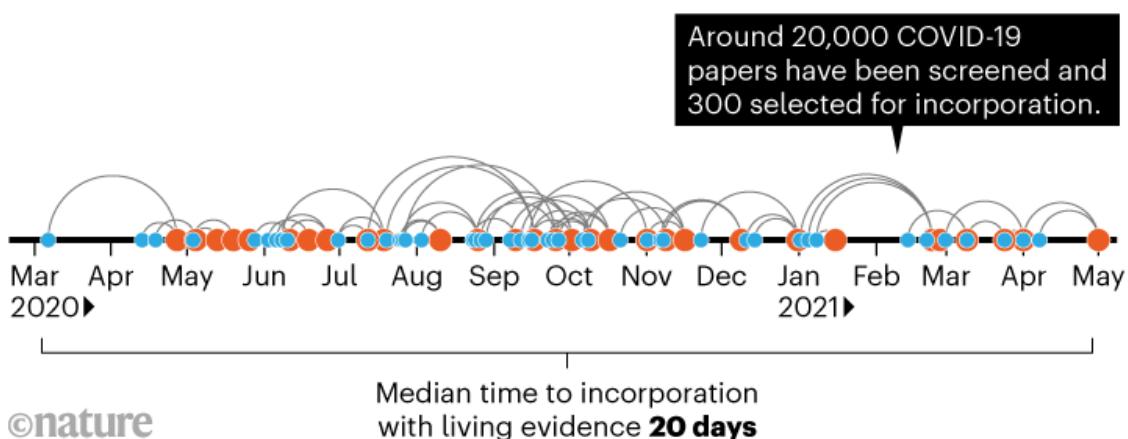
## Stroke

The Australian Stroke Foundation reduced the time between guideline updates from 7 years to under 3 months.



## COVID-19

Learning from the stroke experience, Australian COVID-19 guidelines launched using living evidence, often updating weekly.



Source: Kelvin Hill, Heidi Li, Simon Turner, Jordi Elliott, Andrew Duan

Researchers practising living evidence establish protocols that commit to regular updates and continuously monitor bibliographic and other databases to identify new research. These projects are often enabled by technologies such as natural-language processing, machine learning and crowdsourcing and, sometimes, by FAIR data sharing (which ensures the findability, accessibility, interoperability and reuse of digital assets). Researchers rapidly incorporate studies in frequent cycles that follow established methods for high-quality synthesis and guidelines (for instance, predefined protocols and careful appraisal of the research included).

Even with a deluge of new research, it is feasible to update syntheses and treatment guidelines frequently. For COVID-19, we pushed that limit lower than ever, updating some living guidelines weekly (see ‘Evidence accelerated’). Frequent updating assured policymakers and the public that all relevant research had been taken into account when recommending against the use of the drug hydroxychloroquine, for example. And the COVID-END collaboration used living evidence to keep Canada’s public-health agency updated on COVID-19 vaccine effectiveness as variants of concern emerged.

This year, the UK National Institute for Health and Care Excellence (NICE) announced that living guidelines will be a pillar of its improvement strategy over the next five years. After piloting efforts for several years, the WHO this year began promoting living systematic reviews and living guidelines as standard methodology. It has used the approach for COVID-19, maternal and perinatal care and contraception. In the academic literature, about 100 living systematic reviews have been published since the start of 2020, compared with perhaps 15 in the 4 years previously.



### A fresh approach to evidence synthesis

Many global challenges can benefit from living evidence. Consider climate policy. The Intergovernmental Panel on Climate Change (IPCC) has been very effective in synthesizing evidence on the anthropogenic causes of climate change, its impacts and the potential effects of long-term emission-reduction strategies. But there is no comparable culture of evidence synthesis on the effectiveness of climate-change mitigation or adaptation<sup>3</sup>. Living evidence could help to firmly establish this capacity to ensure that the thousands of decisions needed to achieve net-zero emissions are supported by reliable and up-to-date scientific evidence. There is real urgency for funders and policymakers to incentivize this cultural shift.

Stakeholders and end-users should be involved in both the design and production of living evidence. Depending on the topic, this might mean including communities that are most likely to be affected by extreme weather, or vocational training programmes or a national disease-treatment guideline. For example, early in the pandemic, the COVID-19 Core Outcome Set initiative identified the most important COVID-19 outcomes for patients and clinicians from around the world<sup>4</sup>. This was used to guide how living-guidelines panels judged the potential benefits of new treatments.

Living evidence is particularly relevant for research funders. To direct resources to important questions, funders must identify current knowledge

gaps. Living evidence reveals these gaps by mapping published and ongoing research and by keeping up to date with new research. Investments in living evidence can therefore benefit society twice — directly by funding synthesis efforts, and indirectly by enhancing research investments. Funders should invest in systems, tools and partnerships that will build overall capacities beyond any specific project.

## Four steps

How best can living evidence expand? Four considerations are key.

**Reduce unit costs.** Cutting the time and effort required for systematic review is crucial to the scale-up of living evidence. At present, an initial systematic review typically requires more than 200 person-hours of manual work to sift through titles and abstracts of potentially relevant studies. Many of us are developing technologies to make workflows more efficient and cost-effective. These include natural-language processing and machine learning<sup>5</sup>, crowdsourcing<sup>6</sup>, FAIR data and non-profit software tools, including [Covidience](#) and [EPPI-Reviewer](#) for systematic review, and [MAGICapp](#) for guideline development (J.E. and P.O.V are chief executives at Covidience and MAGICapp, respectively). Together, these technologies could slash the time required by two-thirds. The more others get involved, the better and faster these tools will get. Publications should include better metadata to improve discovery, and should follow FAIR practices to provide open and machine-readable research data.

**Prevent research waste.** As systematic reviews became widely recognized as a form of publication, there was an increase in wasteful, duplicated, low-quality reviews. Some researchers and journals pursue ‘quick’ systematic reviews as a way to boost their publication and citation counts. As with the remdesivir example, the number of systematic reviews summarizing a body of randomized trials can often outnumber the number of relevant trials.



### Four principles to make evidence synthesis more useful for policy

Funders could help by focusing on a set of high-quality living systematic reviews and guidelines. Another approach could be to establish larger collaborative projects, analogous to global multisite clinical trials or ‘big science’ collaborations in physics, as a mechanism to reduce redundancy.

**Optimize publishing.** Publishers such as F1000, Cochrane, *The BMJ* and *Annals of Internal Medicine* have shown that academic publishers can support frequent updating of living systematic reviews and guidelines with multiple updates of essentially the same document. Most are minor and only occasionally major. Workflows were modified so that editorial and peer review, as well as copyediting and production, supported frequent revisions. This required differentiating between small and large alterations and ensuring links between versions. Minor updates can be incorporated in the original publication, or as an addendum. When authors and editors judge that changes to a paper warrant a major update, this can trigger a new article version (linked to previous versions through CrossRef and other links on the journal website) and a new bibliographic database listing and digital object identifier (DOI).

The research community must rethink what the ‘version of record’ means and how to optimize the incentives and rewards for authors and peer reviewers of frequently updated systematic reviews. Peer review of these

continuously updated documents should be neither slow and onerous nor a quick rubber stamp. The underlying data from living systematic reviews and guidelines should be FAIR, to support replicability and transparency and maximize reuse.

**Facilitate implementation.** Living evidence has the potential to speed up the incorporation of science into practice and policy. Decisions do not depend solely on evidence, however, and changes in practice generally lag behind changes in guidance. Research on implementation is therefore essential to understand how best living evidence can serve policy and practice. So far, experience with implementation has been limited to fewer than a dozen case studies such as those mentioned here.



### [How COVID broke the evidence pipeline](#)

Issues that need to be investigated include: assessing the impact of new models of dissemination (for example, continuous ‘what’s new’ feeds rather than conventional, intermittent announcements); deciding when changes in evidence should trigger implementation; and whether a supply of reliable, current evidence affects how it is incorporated into decisions. In many instances, implementation activities will not need to change, and living evidence will give them a head start.

Living-evidence outputs should also be set up to integrate with data systems. For example, health-care recommendations in living guidelines could flow into clinical tools to support physicians' decisions. Living systematic reviews of low-carbon technologies could update climate-change-mitigation models.

## Keep moving

Decisions relevant to global challenges must be informed by the best available evidence. It should no longer be acceptable for evidence to be out of date, biased or selective. Without trustworthy and up-to-date summaries, the world risks making ill-informed decisions and wasting investment.

Living evidence as a practice has been around for barely five years, and there is much to learn. Our experience so far is counter to common initial concerns. Rather than being confused by changing guidelines, clinicians value a resource they know is up to date. Interest from living-evidence practitioners can be sustained and team membership can gradually shift without interrupting updates. The cost of continuously updating guidelines seems about the same as gearing up for big revisions every few years.

The advances made in evidence systems during the COVID-19 pandemic should extend beyond health. We call on researchers in all scientific fields, and their funders, to test the living-evidence model across diverse domains. Trialling the approach with different types of evidence, for a wide range of decision makers, will contribute to the advancement and scale-up of the model. Science does not stand still; neither should its synthesis and translation into action.

*Nature* **600**, 383-385 (2021)

doi: <https://doi.org/10.1038/d41586-021-03690-1>

The contents of this article are the responsibility of the authors and do not reflect the views of the World Health Organization, for which O.T.O. works.

## References

1. 1.

*Nature* **593**, 182–185 (2021).

2. 2.

Elliott, J. H. *et al.* *PLoS Med* **11**, e1001603 (2014).

3. 3.

Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J. & Edenhofer, O. *Environ. Sci. Policy* **77**, 252–259 (2017).

4. 4.

Tong, A. *et al.* *Crit. Care Med.* **48**, 1622–1635 (2020).

5. 5.

Thomas, J. *et al.* *J. Clin. Epidemiol.* **133**, 140–151 (2021).

6. 6.

Noel-Storr, A. *et al.* *J. Clin. Epidemiol.* **133**, 130–139 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03690-1>

- CORRESPONDENCE
- 10 December 2021

# To beat Omicron, Delta and bird flu, Europe must pull together

- [Amélie Desvars-Larrive](#)    ORCID: <http://orcid.org/0000-0001-7671-696X> <sup>0</sup>

Omicron, the newly discovered variant of SARS-CoV-2, is compounding concerns in European and other countries, many of which are already experiencing a surge in infections caused by the Delta variant. Meanwhile, outbreaks of highly pathogenic avian influenza in poultry and wild birds are intensifying in Eurasia and Africa, keeping epidemiologists on high alert because of the zoonotic and pandemic potential of the H5Nx viruses responsible.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 386 (2021)

doi: <https://doi.org/10.1038/d41586-021-03661-6>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03661-6>

| [Section menu](#) | [Main menu](#) |

- CORRESPONDENCE
- 10 December 2021

# Chile: elect a president to strengthen climate action, not weaken it

- [Maisa Rojas](#) ORCID: <http://orcid.org/0000-0002-5985-1905><sup>0</sup>,
- [Juan Carlos Muñoz](#) ORCID: <http://orcid.org/0000-0003-1775-985X><sup>1</sup>,
- [Rodrigo Palma-Behnke](#)<sup>2</sup>,
- [Pablo A. Marquet](#) ORCID: <http://orcid.org/0000-0001-6369-9339><sup>3</sup>,
- [Alejandra Stehr](#)<sup>4</sup> &
- [Humberto González](#) ORCID: <http://orcid.org/0000-0002-3588-3451><sup>5</sup>

As leading climate scientists in Chile, we are alarmed that a candidate who openly questions the current human influence on climate change became a front runner in the first round of the country's presidential elections last month.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.  
VAT will be added later in the checkout.  
Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 386 (2021)

doi: <https://doi.org/10.1038/d41586-021-03662-5>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03662-5>

| [Section menu](#) | [Main menu](#) |

- CORRESPONDENCE
- 14 December 2021

# Prepare PhD holders for different career tracks

- [Theo van den Broek](#)<sup>0</sup>,
- [Berent Prakken](#)<sup>1</sup> &
- [Frank Miedema](#)<sup>2</sup>

Many of the graduate students who aim to become principal investigators in academic research laboratories are likely to be disappointed. Exposing them to other options by means of networking and internships can help to guide them to success in more diverse careers (see, for example, [E. Juskewitz et al. \*Nature\* <https://doi.org/g8nb>; 2021](#)). They need to be made aware early in their training that they can pursue alternative fulfilling careers, inside or outside academia.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 386 (2021)

doi: <https://doi.org/10.1038/d41586-021-03695-w>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03695-w>

| [Section menu](#) | [Main menu](#) |

- CORRESPONDENCE
- 14 December 2021

# Remembering a pioneer in biotechnology

- [Tzvi Aviv](#) <sup>0</sup>,
- [Frank Sicheri](#) <sup>1</sup>,
- [Alan Bernstein](#) <sup>2</sup> &
- [Robert A. Weinberg](#) <sup>3</sup>

Haim Aviv, inventor of a technique that paved the way for molecular biology applications in agriculture and pharmaceuticals, died last month in Rehovot, Israel. Born in Arad, Romania, in 1940, he worked on messenger RNA for his PhD at Israel's Weizmann Institute of Science.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 386 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-03696-9>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03696-9>

| [Section menu](#) | [Main menu](#) |

# Work

- **[Trapped in a hotel room: my scientific life in the pandemic](#)**

[ 26 November 2021]

Career Column • Jen Lewendon's move from the United Kingdom for a postdoc restricted her travel and led to extended stints in quarantine. Here's what the experience taught her.

- **[Something in the air: gathering dust that's crossed an ocean](#)** [ 13 December 2021]

Where I Work • Edmund Blades manages an atmospheric observatory at Ragged Point, Barbados.

- CAREER COLUMN
- 26 November 2021

# Trapped in a hotel room: my scientific life in the pandemic

Jen Lewendon's move from the United Kingdom for a postdoc restricted her travel and led to extended stints in quarantine. Here's what the experience taught her.

- [Jen Lewendon](#) 0



Each time Jen Lewendon flew into Hong Kong International Airport, pictured here in May 2020, she had to quarantine for 21 days. Credit: Tommy Walker/NurPhoto/Getty

I've spent more than two of the past five months self-isolating in hotel rooms in Thailand and Hong Kong, forbidden to step outside. During these

stints I have, at times, stood before a mirror just to appreciate an unmasked human face staring back at me. I've looked forward to the knock on the door indicating that a stranger, dressed from head to toe in personal protective equipment, is waiting to stick a swab up my nose and down my throat. I've stared through the peep-hole, excited by the warped, goldfish view of a hotel assistant dropping my meal off in the corridor. And I've lain in bed at 10 p.m. watching some creepy thriller, only to jump out of my skin as, without warning, someone wipes down my door with disinfectant.

Travel restrictions in Hong Kong stipulate mandatory quarantines of up to 21 days on arrival, and, in some cases, a further 21 days in a lower-risk country (in my case, Thailand) before entry. For most, these periods spent alone in small, sound-insulated rooms are a bizarre departure from life as we normally know it. Nonetheless, I consider myself lucky. After finishing my PhD in cognitive neuroscience at Bangor University, UK, in 2020, I faced the career abyss experienced by many early-career researchers during the COVID-19 pandemic. Graduate students all around me were having to pause their studies, or in some cases drop out completely, because their funding necessitated research that they could no longer conduct.



### [Collection: How to move lab](#)

In 2020, I secured a postdoc at the Hong Kong Polytechnic University after more than a year of writing job applications. Straight after my PhD, I'd been

offered a job in Singapore. I'd applied for the visa and had packed my bags when the pandemic hit. My prospective employer — presumably affected by funding cuts — stopped returning my e-mails. So, when this job in Hong Kong came up, I jumped at the opportunity to work with a fantastic principal investigator on a project studying the electrophysiological correlates of speech processing. The fact that Hong Kong had some of the most stringent COVID-19 restrictions in the world was definitely an after-thought.

Facing the true extent of travel-restriction complexities is now very much a reality for me, as it is for many other researchers, with home and work life split between countries that have entirely different approaches to COVID-19.

Here are some lessons I've learnt so far.

## **Manage work and social life in isolation**

Isolation will be different for everyone, but I found that there are a couple of things you can do to maximize your chances of surviving quarantine while retaining a strand of sanity. First, make the most of the opportunity to catch up with people. Respond to old unread messages and take the time to get in touch with friends. And laboratory meetings might just become your only source of contact (albeit virtual) during working hours, so take advantage of the social input they provide.

Keep moving, keep eating — and don't let work take over. It took me a while to realize that my step count should probably be higher than my word count, even while stuck inside. Allow your calorie count to be an outlier — food can be a great source of joy and structure in an otherwise very limited world. Mere restrictions on movement weren't going to hinder the culinary tour of Hong Kong that I'd embarked on from the confines of my quarantine hotel room, through delivery services.



### [Collection: Young scientists](#)

Finally, set work and non-work goals, if these are helpful, but be kind to yourself if (when) you don't meet them. Some days, it might prove difficult enough to muster the enthusiasm to get out of bed in the morning, let alone tackle a particularly dull work task that you've been putting off for weeks.

## **Set your non-negotiables**

If you're heading to a country with strict travel restrictions and expensive entry and testing procedures, ask yourself some key questions and try to answer honestly. How often do you want to be able to go home and see your partner, family or friends? How much are you willing to spend on tests and quarantines so that you can leave and return to that country? How comfortable are you with the idea of total isolation, and how long could you put up with it? Establish what you're happy with and write it down, so as to have a concrete baseline to refer back to, because in the ever-changing world of travel restrictions and requirements, you can bet that your non-negotiables are going to be challenged.



Hong Kong's stringent COVID-19 restrictions meant that Jen Lewendon had to spend long periods confined to her hotel room. Credit: Jennifer Lewendon

I knew that, as for many language scientists, data collection for my research would require extended trips outside Hong Kong, and I appreciated the opportunity — through collaborations in Wales and Ireland — to be closer to home for periods during my work. Nonetheless, after two trips between Hong Kong and the United Kingdom, one involving 8 weeks of isolation, a ‘maximum 21 days quarantine’ rule is the latest addition to my non-negotiable list.

## **Push for flexibility**

We're all living through unprecedeted times, and you shouldn't be the only one making the compromise. Don't be afraid to explore possibilities with

your employer in terms of remote work, flexible working arrangements and collaborations that take you a little closer to home. If prospective employers are entirely unwilling to engage in negotiation, take some time to consider whether they're the right people to be employed by during a global pandemic.

## Be realistic

COVID tests, quarantines, more COVID tests, cancelled flights, ‘fit to fly’ certificates and other realities of travelling in the pandemic mean that, if you’re thinking of committing to a job in a high-restriction country that requires travel, or have any desire to visit home, it’s going to cost you. It’s possible that it’s not even going to be a wise decision, financially speaking, to take the job in the first place. For me, the financial security of my first postdoc, after nearly a decade of studying, has been somewhat offset by the travel protocols constraining a research project that necessitates international travel, and by the difficulty of returning home and maintaining a work–life balance. When considering a position, ask yourself: How much do you love this project? Are you purely motivated by the research and opportunities? Because, as an early-career researcher, if you’re doing it for the money, it might not pay off.

## Act if you’re not happy

Ever-varying travel restrictions and entry requirements can sometimes feel like a bad relationship, in which the goalposts are continually being moved and the compromises you’ve made are constantly challenged.

For me, being denied boarding at the airport because of a minor discrepancy on a document (after immigration had confirmed that it was OK), and waiting until 5 a.m. for the ‘misplaced’ results of a COVID-19 test (mere hours before a flight) were real low points. Keep checking with yourself whether you’ve crossed the threshold, and the situation is no longer working for you. It’s common in academia to accept being less than happy for brief stints, so as to make progress, but remember that being unhappy shouldn’t be the norm.

*Nature* **600**, 559-560 (2021)

doi: <https://doi.org/10.1038/d41586-021-03566-4>

This is an article from the Nature Careers Community, a place for Nature readers to share their professional experiences and advice. [Guest posts are encouraged.](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03566-4>

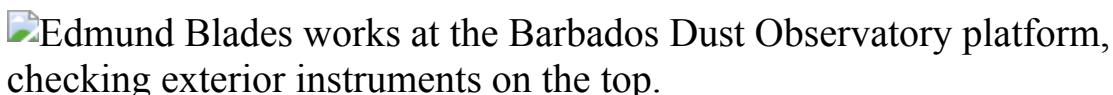
| [Section menu](#) | [Main menu](#) |

- WHERE I WORK
- 13 December 2021

# Something in the air: gathering dust that's crossed an ocean

Edmund Blades manages an atmospheric observatory at Ragged Point, Barbados.

- [Amber Dance](#) <sup>0</sup>



Edmund Blades is site manager at the University of Miami's Barbados Atmospheric Chemistry Observatory at Ragged Point, and a microbiologist at Queen Elizabeth Hospital in Bridgetown, Barbados. Credit: Micah B. Rubin for *Nature*

Tourists come here to Ragged Point, at the easternmost part of Barbados, for the view. I climb the 17-metre tower of the University of Miami's Barbados Atmospheric Chemistry Observatory and collect samples of ash, dust and microorganisms, some of which have come all the way across the Atlantic Ocean. Reddish, iron-rich clouds from Africa reach us every year, usually between June and September, making Barbados hazy. When it's clear, I enjoy the air and the landscape while making scientific observations — behind me in the photo is the northern coast.

The observatory has been collecting dust data since 1966. As site manager, I liaise with collaborators and lead tours for local students. I earned my PhD here, from 1996 to 2005, investigating the microbes that reach Barbados from Africa and their potential health effects. Many people suspected that the dust caused asthma attacks.

Nearly every day for those ten years, I climbed the tower to collect filters in pumps like the one I'm working on here. I took the filters back to the Best-dos Santos Public Health Laboratory in Bridgetown, which I was managing at the time, to grow bacteria and fungi. I was so surprised to find living microbes; I didn't think they'd survive the transatlantic crossing. These were spore-formers — microbes that can form a tough coating to withstand harsh conditions, but that don't usually affect people.

I found no correlation between dusty days and visits by children to the asthma clinic at the emergency department of Bridgetown's Queen Elizabeth Hospital. Rather, asthma visits correlated with local grass pollen and rainfall. This research led to a further surprise in 2020, when I received the Barbados Service Star, a medal from the government, in recognition of my contribution to public health.

*Nature* **600**, 562 (2021)

doi: <https://doi.org/10.1038/d41586-021-03706-w>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03706-w>

| [Section menu](#) | [Main menu](#) |

# Research

- **[Context is key for learning motor skills](#)** [ 24 November 2021]  
News & Views • A sophisticated theory for learning motor skills places emphasis on the need for inferring context — drawing conclusions about the structure of the environment — for efficiently storing and expressing motor memories.
- **[Hominin footprints at Laetoli reveal a walk on the wild side](#)** [ 01 December 2021]  
News & Views • Bipedalism is a defining feature of the human lineage, but not all hominin species walked in the same way. New data from a famous palaeoanthropology site reveal that at least two differently bipedal hominins roamed eastern Africa.
- **[Human genetic variants identified that affect COVID susceptibility and severity](#)** [ 08 July 2021]  
News & Views • An individual's genetics can influence their risk of infection and the severity of disease symptoms. A large international study has identified parts of the human genome that can affect the risk of severe COVID-19.
- **[Benefits of megastudies for testing behavioural interventions](#)** [ 08 December 2021]  
News & Views • Trials of behavioural interventions are hard to compare, hampering policy decision-making. The effects of more than 50 interventions on exercise behaviour have been compared using an experimental design called a megastudy.
- **[The emergence, genomic diversity and global spread of SARS-CoV-2](#)** [ 08 December 2021]  
Review Article • The potential origins and global spread of SARS-CoV-2, the emergence of SARS-CoV-2 variants of concern and the importance of genomic surveillance for the control of the COVID-19 pandemic are discussed.
- **[Biogeochemical extremes and compound events in the ocean](#)** [ 15 December 2021]  
Perspective • High-temperature, high-acidity and low-oxygen extremes may pose a particular threat to marine ecosystems, requiring a major effort to understand them and the ability of marine life to respond to them.
- **[Sublimation-driven convection in Sputnik Planitia on Pluto](#)**  
[ 15 December 2021]  
Article • A modelling study describing the formation of the polygonal surface structures in Sputnik Planitia on Pluto shows that convection driven by ice sublimation can generate

planetary-scale surface patterns.

- **Direct limits for scalar field dark matter from a gravitational-wave detector** [ 15 December 2021]  
Article • Using a gravitational-wave detector to listen for dark matter signatures, a direct search for scalar field dark matter was conducted and new upper limits are set on the coupling constants.
- **Observation of Feshbach resonances between a single ion and ultracold atoms** [ 15 December 2021]  
Article • Magnetically tunable interactions between lithium atoms and barium ions are used to demonstrate and probe Feshbach resonances between atoms and ions, which could have applications in the fields of experimental quantum simulation and fundamental physics.
- **Detecting spins by their fluorescence with a microwave photon counter** [ 15 December 2021]  
Article • An ensemble of electron spins is detected by their microwave fluorescence using a superconducting single microwave photon counter, making single-spin electron spin resonance spectroscopy a possible future prospect.
- **Fractional Chern insulators in magic-angle twisted bilayer graphene** [ 15 December 2021]  
Article • A study using local compressibility measurements reports fractional Chern insulator states at low magnetic field in magic-angle twisted bilayer graphene, and establishes the applied magnetic field as a means to tune the Berry curvature distribution.
- **Tritiation of aryl thianthrenium salts with a molecular palladium catalyst** [ 15 December 2021]  
Article • The isotopic label tritium can be selectively added into aromatic organic compounds by a homogenous hydrogenolysis reaction using aryl thianthrenium salts, tritium gas and a molecular palladium catalyst.
- **A large West Antarctic Ice Sheet explains early Neogene sea-level amplitude** [ 15 December 2021]  
Article • Variations in Miocene sea level can be explained by a large marine-based West Antarctic Ice Sheet.
- **Uncovering global-scale risks from commercial chemicals in air** [ 15 December 2021]  
Article • A new framework is proposed for assessing the risks of the atmospheric transformation products of commercial chemicals, combining laboratory and field experiments, advanced techniques for screening suspect chemicals, and in silico modelling.

- **Deep-mantle krypton reveals Earth's early accretion of carbonaceous matter** [ 15 December 2021]  
Article • The krypton isotopic pattern of Earth's deep mantle indicates that volatile-rich material from the outer Solar System was delivered early in Earth's accretion history.
- **Footprint evidence of early hominin locomotor diversity at Laetoli, Tanzania** [ 01 December 2021]  
Article • Reanalysis of bipedal trackways from Laetoli site A in Tanzania suggest that the footprints were made by a hominin that coexisted with at least one other hominin species.
- **Mapping the human genetic architecture of COVID-19** [ 08 July 2021]  
Article • A global network of researchers was formed to investigate the role of human genetics in SARS-CoV-2 infection and COVID-19 severity; this paper reports 13 genome-wide significant loci and potentially actionable mechanisms in response to infection.
- **Megastudies improve the impact of applied behavioural science** [ 08 December 2021]  
Article • A massive field study whereby many different treatments are tested synchronously in one large sample using a common objectively measured outcome, termed a megastudy, was performed to examine the ability of interventions to increase gym attendance by American adults.
- **Cognitive control persistently enhances hippocampal information processing** [ 10 November 2021]  
Article • Studies in mice show that cognitive control training rapidly improves brain circuit function and enhances subsequent learning, which both persist for months.
- **Contextual inference underlies the learning of sensorimotor repertoires** [ 24 November 2021]  
Article • A theory of motor learning based on the principle of contextual inference reveals that adaptation can arise by both creating and updating memories and changing how existing memories are differentially expressed, and predicts evoked recovery and context-dependent single-trial learning.
- **Exercise plasma boosts memory and dampens brain inflammation via clusterin** [ 08 December 2021]  
Article • Plasma from voluntarily running mice reduces baseline expression of neuroinflammatory genes and experimentally induced brain inflammation when infused into sedentary mice.
- **Combinatorial, additive and dose-dependent drug–microbiome associations** [ 08 December 2021]

Article • An analysis of 2,173 individuals from the MetaCardis cohort quantifies the individual and combinatorial effects of a range of drugs on host health, metabolome and gut microbiome in cardiometabolic disease.

- **Genomic reconstruction of the SARS-CoV-2 epidemic in England** [ 14 October 2021]

Article • A study of the evolution of the SARS-CoV-2 virus in England between September 2020 and June 2021 finds that interventions capable of containing previous variants were insufficient to stop the more transmissible Alpha and Delta variants.

- **High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape** [ 20 September 2021]

Article • A complex range of mutations within the SARS-CoV-2 spike protein is needed to escape polyclonal plasma neutralizing antibodies, and plasma from individuals who were first infected then vaccinated display the greatest resilience to escape mutations.

- **Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination** [ 07 October 2021]

Article • Individual memory antibodies selected over time by natural infection with SARS-CoV-2 have greater potency and breadth than antibodies elicited by vaccination, whereas the overall neutralizing potency of plasma is greater following vaccination.

- **Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity** [ 11 October 2021]

Article • Immune analyses against circulating SARS-CoV-2 variants show that mRNA vaccination induces robust neutralizing antibodies and boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection.

- **Hybrid immunity improves B cells and antibodies against SARS-CoV-2 variants** [ 20 October 2021]

Article • Single-cell-level analysis of memory B cells and their response to vaccination against all SARS-CoV-2 variants of concern in individuals who either had or had not been previously exposed to the virus.

- **A multi-scale map of cell structure fusing protein images and interactions** [ 24 November 2021]

Article • Protein immunofluorescence imaging and affinity purification–mass spectrometry are combined to create a unified map of human cell architecture across scales, which the authors call the multi-scale integrated cell (MuSIC).

- **Accuracy mechanism of eukaryotic ribosome translocation**

[ 01 December 2021]

Article • Structural analysis of the *Saccharomyces cerevisiae* 80S ribosome trapped in an intermediate translocation state shows stabilization of codon–anticodon interactions by eukaryote-specific elements of the 80S ribosome, eEF2 and tRNA and demonstrates a major role for eEF2 in maintaining the directionality of translocation.

- **De novo protein design by deep network hallucination** [ 01

December 2021]

Article • The trRosetta neural network was used to iteratively optimise model proteins from random 100-amino-acid sequences, resulting in ‘hallucinated’ proteins, which when expressed in bacteria closely resembled the model structures.

- **The conformational cycle of prestin underlies outer-hair cell electromotility** [ 25 October 2021]

Article • Mammalian prestin underlies cellular electromotility by sensing the voltage through a dynamic sensor and converting it to an in-plane area expansion.

- **On the liquid–liquid phase transition of dense hydrogen**

[ 15 December 2021]

Matters Arising •

- **Reply to: On the liquid–liquid phase transition of dense hydrogen** [ 15 December 2021]

Matters Arising •

- NEWS AND VIEWS
- 24 November 2021

# Context is key for learning motor skills

A sophisticated theory for learning motor skills places emphasis on the need for inferring context — drawing conclusions about the structure of the environment — for efficiently storing and expressing motor memories.

- [Anne G. E. Collins](#)<sup>0</sup> &
- [Samuel D. McDougle](#)<sup>1</sup>

Humans can develop a wide variety of motor skills, from speaking and riding bicycles, to manipulating complex objects such as computer keyboards and musical instruments. Having a robust portfolio of motor skills encoded as motor memories is crucial to everyday life, but the neural and psychological principles of motor learning and memory are in many ways still a mystery. [Writing in Nature](#), Heald *et al.*<sup>1</sup> examine the computational principles of motor-memory formation and modification. By placing context centre stage in motor learning, the authors account for various behavioural phenomena in a single, elegant computational framework.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.  
VAT will be added later in the checkout.  
Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

## Rent or Buy

All prices are NET prices.

## **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 387–388 (2021)

doi: <https://doi.org/10.1038/d41586-021-03028-x>

## **References**

1. 1.

Heald, J. B., Lengyel, M. & Wolpert, D. M. *Nature* **600**, 489–493 (2021).

2. 2.

Thoroughman, K. A. & Shadmehr, R. *Nature* **407**, 742–747 (2000).

3. 3.

Kitago, T., Ryan, S. L., Mazzoni, P., Krakauer, J. W. & Haith, A. M. *Front. Hum. Neurosci.* **7**, 307 (2013).

4. 4.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S. & Wyart, V. *Nature Neurosci.* **22**, 2066–2077 (2019).

5. 5.

Collins, A. & Koechlin, E. *PLoS Biol.* **10**, e1001293 (2012).

6. 6.

Gershman, S. J., Radulescu, A., Norman, K. A. & Niv, Y. *PLoS Comput. Biol.* **10**, e1003939 (2014).

7. 7.

Collins, A. G. E. *J. Cogn. Neurosci.* **30**, 1422–1432 (2018).

8. 8.

McDougle, S. D., Bond, K. M. & Taylor, J. A. *J. Neurophysiol.* **118**, 383–393 (2017).

9. 9.

Krakauer, J. W., Hadjiosif, A. M., Xu, J. , Wong, A. L. & Haith, A. M. *Compreh. Physiol.* **9**, 613–663 (2019).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03028-x>

- NEWS AND VIEWS
- 01 December 2021

# Hominin footprints at Laetoli reveal a walk on the wild side

Bipedalism is a defining feature of the human lineage, but not all hominin species walked in the same way. New data from a famous palaeoanthropology site reveal that at least two differently bipedal hominins roamed eastern Africa.

- [Stephanie M. Melillo](#) 

The human version of walking on two legs, known as striding bipedalism, is unique among mammals. It requires the ability to balance a tower of loosely connected body parts over a single foot, as the other foot swings forwards to complete the stride. Conventional wisdom holds that this ungainly form of locomotion had a single evolutionary origin in an ancestral hominin, followed by about six million years during which further anatomical adjustments accumulated — a linear model of evolution in which early hominin bipedalism became progressively more similar to our own over time. However, fossils discovered during the past decade show that multiple versions of bipedalism existed simultaneously during one or more periods of hominin evolution. [Writing in Nature](#), McNutt *et al.*<sup>1</sup> suggest that evidence of locomotor diversity in hominins has been overlooked for many decades.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 388-390 (2021)

doi: <https://doi.org/10.1038/d41586-021-03469-4>

## **References**

1. 1.

McNutt, E. J. *et al.* *Nature* **600**, 468–471 (2021).

2. 2.

Leakey, M. D. & Harris, J. M. *Laetoli. A Pliocene Site in Northern Tanzania* (Clarendon, 1987).

3. 3.

Tuttle, R. H., Webb, D. M. & Tuttle, N. I. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 187–198 (CNRS, 1991).

4. 4.

Tuttle, R. H., Webb, D. M., Tuttle, N. I. & Baksh, M. in *Topics in Primatology* Vol. 3 (eds Matano, S., Tuttle, R. H., Ishida, H. & Goodman, M.) 221–242 (Univ. Tokyo Press, 1992).

5. 5.

Wood, B. & Boyle, E. K. *Am. J. Phys. Anthropol.* **159**, S37–S78 (2016).

6. 6.

Kimbrel, W. H. in *Handbook of Paleoanthropology* (eds Henke, W. & Tattersall, I.) 2071–2105 (Springer, 2015).

7. 7.

Haile-Selassie, Y. *et al. Nature* **483**, 565–569 (2012).

8. 8.

Lovejoy, C. O., Latimer, B., Suwa, G., Asfaw, B. & White, T. D. *Science* **326**, 72 (2009).

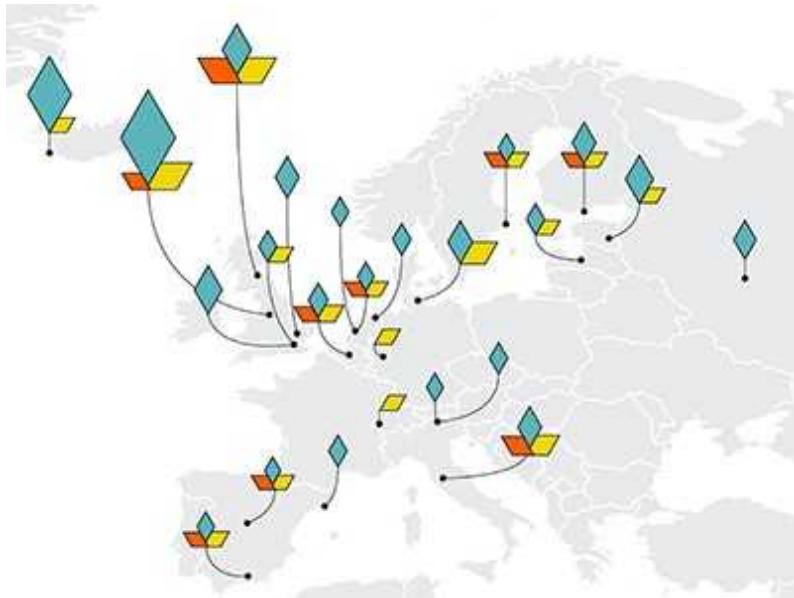
- NEWS AND VIEWS
- 08 July 2021

# Human genetic variants identified that affect COVID susceptibility and severity

An individual's genetics can influence their risk of infection and the severity of disease symptoms. A large international study has identified parts of the human genome that can affect the risk of severe COVID-19.

- [Samira Asgari](#) ORCID: <http://orcid.org/0000-0002-2347-8985><sup>0</sup> &
- [Lionel A. Pousaz](#)<sup>1</sup>

For more than a year now, scientists and clinicians have been trying to understand why some people develop severe COVID-19 whereas others barely show any symptoms. Risk factors such as age and underlying medical conditions<sup>1</sup>, and environmental factors including socio-economic determinants of health<sup>2</sup>, are known to have roles in determining disease severity. However, variations in the human genome are a less-investigated source of variability. [Writing in Nature](#), members of the COVID-19 Host Genetics Initiative<sup>3</sup> ([www.covid19hg.org](http://www.covid19hg.org)) report results of a large human genetic study of SARS-CoV-2 infection. The researchers identify 13 locations (or loci) in the human genome that affect COVID-19 susceptibility and severity.



### [Read the paper: Mapping the human genetic architecture of COVID-19](#)

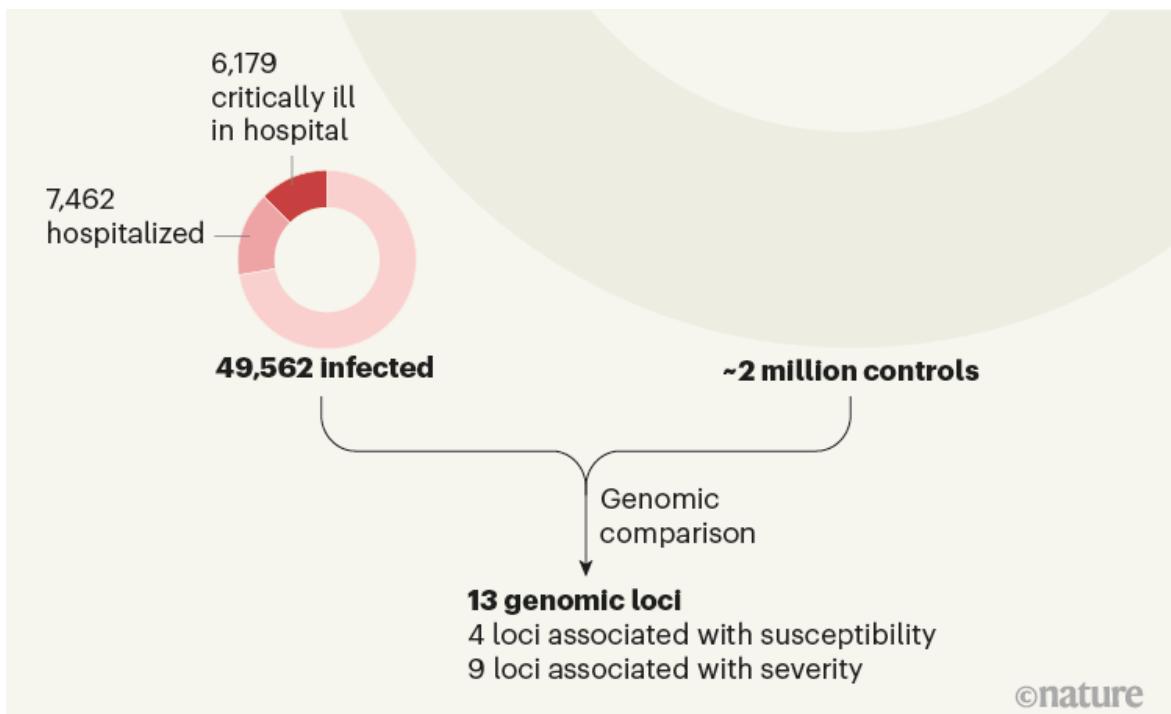
Scientists already knew that human genetic variants can influence the severity of infectious diseases, including infection with SARS-CoV-2<sup>4-6</sup>. The effects of genetic factors range from those of rare, high-impact mutations that can make the difference between an individual developing mild symptoms and life-threatening illness<sup>7</sup>, to more-common genetic variants that only moderately affect symptom severity<sup>5</sup>.

Even so, human genomic studies of infectious diseases remain scarce compared with those of other immune-mediated conditions, such as autoimmune disorders. There are several reasons for that. Chief among them is that infectious diseases are typically studied with a focus on the disease-causing microorganism, rather than the host. Moreover, human genetic variants usually have relatively small effects on infection outcomes compared with the effects of socio-demographic factors such as age or access to health care<sup>8</sup>. Identifying these generally modest effects requires studies of large, well-characterized groups of people to produce sufficient statistical power to reveal the relevant genetic factors. Finally, unlike for chronic diseases, the window for characterizing the severity and outcomes of infectious diseases is often limited to a short period during which individuals are symptomatic.

The authors overcame these challenges by rapidly setting up a large, international collaboration when the pandemic started. This collaboration of around 3,000 researchers and clinicians includes data from 46 studies involving more than 49,000 individuals with COVID-19 and 2 million control individuals, with participants recruited from 6 ancestry groups and 19 countries. By acting swiftly, the authors could recruit symptomatic patients, and, by setting up international collaborations, were able to include enough participants to overcome statistical-power limitations. In addition, they tried to account for the role of socio-demographic factors by collecting data on some of the known risk factors, such as age and sex, and including this information in their statistical analyses.

To obtain comparable results across all 46 study groups, the authors defined 3 categories of analysis: infection, which included people with physician-confirmed, laboratory-confirmed or self-reported COVID-19; hospitalization, which consisted of individuals with laboratory-confirmed moderate to severe COVID-19; and critical illness, patients with laboratory-confirmed infection who were hospitalized and required respiratory support or died. To identify genetic variants associated with COVID-19 susceptibility and severity, the authors first compared the difference in the frequencies of millions of genetic variants between the people infected with COVID-19 and the control individuals in each study. They then combined the results from all 46 studies to increase the statistical power of their data.

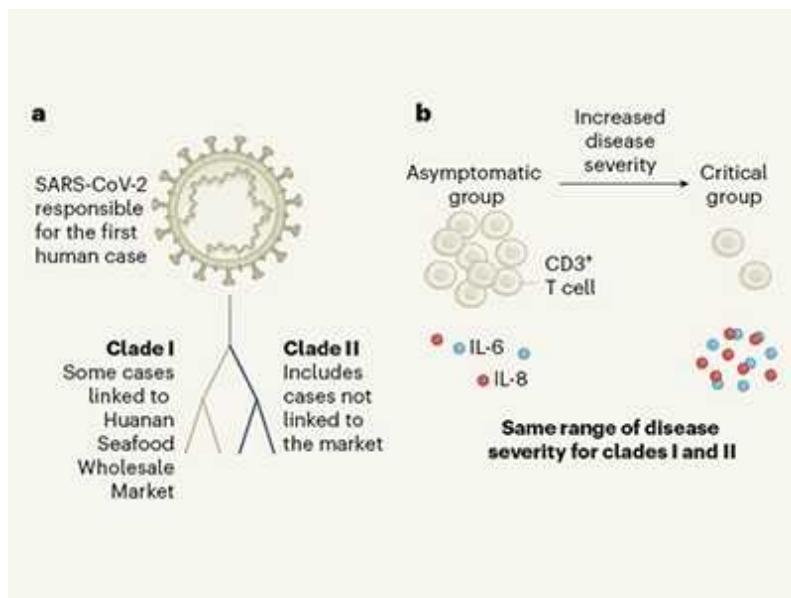
Through this combined analysis, the authors identified 13 loci that were associated with SARS-CoV-2 infection and disease severity (Fig. 1), including 6 loci not reported in previous human genomics studies of COVID-19<sup>4,5</sup>. Four loci affect general susceptibility to SARS-CoV-2, whereas nine were associated with disease severity. Two of the previously unassociated loci were discovered only when individuals with East Asian ancestry were included in the analysis, highlighting the value of including diverse populations in human genomics studies.



**Figure 1 | Identifying regions of the human genome associated with COVID-19 susceptibility and severity.** The COVID-19 Host Genetics Initiative<sup>3</sup> sought to identify genetic variants that account for the variability in individuals' susceptibility to COVID-19, as well as in the severity of the disease. The authors compared the genomes of 49,562 individuals with COVID-19 (including 13,641 individuals who were hospitalized with the infection and, of those, 6,179 who were critically ill with the disease) with the genomes of around 2 million control individuals without known infection. This comparison pointed to 13 locations in the genome (loci): variants in 4 of these loci are associated with susceptibility to COVID-19, whereas variants in 9 others are associated with disease severity.

To better understand the biology of COVID-19 and the mechanisms that connect these loci to disease outcomes, the authors looked for genes that were in the proximity of each locus (that is, 'candidate genes'). They identified more than 40 candidate genes, several of which have previously been implicated in immune function or have known functions in the lungs, suggesting that variants in the genomic regions highlighted by the authors' findings might exert their effect on COVID-19 outcome through the respiratory system.

One such example is the gene *TYK2*. Variants of this gene can increase susceptibility to infections by other viruses, bacteria and fungi<sup>9</sup>. In line with this, the authors reported that individuals who carry certain mutations in *TYK2* are at increased risk of being hospitalized or developing critical illness from infection with SARS-CoV-2. Another example is the gene *DPP9*. The authors found a variant in this gene that increases the risk of becoming critically ill with COVID-19. Notably, the same variant can increase the risk of a rare pulmonary disease characterized by scarring of the lung tissue<sup>10</sup>.



### A race to determine what drives COVID-19 severity.

This study by the COVID-19 Host Genetics Initiative represents a major milestone in our understanding of the role of human genetics in susceptibility to SARS-CoV-2; however, more work remains to be done. Future experiments should determine all the genes, signalling pathways and biological mechanisms that connect the genomic loci identified to COVID-19 outcomes.

Moreover, despite the authors' efforts to include genetically diverse study groups, about 80% of the participants are of European ancestry. Future studies containing a larger number of individuals from other ancestry groups are needed to ensure that the results apply to non-Europeans, and to identify other loci that might be associated with risk in people of other ancestries.

Another complex question that could not be addressed in the authors' study is the combined effect of specific variants in the SARS-CoV-2 genome and variants in the human genome on disease outcome. Finally, as the authors mention, they could not fully control for all socio-demographic factors, such as access to health care. Although such non-genetic factors are unlikely to explain all the findings, they could bias some of the associations between genetic variants and disease outcome.

Despite these limitations, the implications of the study's results are far-reaching. This study is important not only for advancing our understanding of human susceptibility to COVID-19; it also underlines the value of global collaborations for clarifying the human genetic basis of variability in susceptibility to infectious diseases. Infections remain among the top causes of mortality in lower-income countries, and represent a growing global threat, owing to climate change, urbanization and rising population size<sup>11</sup>. Human genomics can be an effective tool with which to understand the biological mechanisms that underlie immune responses to specific infections, to identify at-risk individuals and to develop new drugs and vaccines for existing or emerging infections.

*Nature* **600**, 390-391 (2021)

doi: <https://doi.org/10.1038/d41586-021-01773-7>

## References

1. 1.

Williamson, E. J. *et al.* *Nature* **584**, 430–436 (2020).

2. 2.

Wang, Z. & Tang, K. *Nature Med.* **26**, 458 (2020).

3. 3.

COVID-19 Host Genetics Initiative. *Nature* **600**, 472–477 (2021)

4. 4.

Severe Covid-19 GWAS Group *et al.* *N. Engl. J. Med.* **383**, 1522–1534 (2020).

5. 5.

Pairo-Castineira, E. *et al.* *Nature* **591**, 92–98 (2021).

6. 6.

Casanova, J.-L., Su, H. C. & COVID Human Genetic Effort. *Cell* **181**, 1194–1199 (2020).

7. 7.

Zhang, Q. *et al.* *Science* **370**, eabd4570 (2020).

8. 8.

Webb Hooper, M., Nápoles, A. M. & Pérez-Stable, E. J. *JAMA* **323**, 2466–2467 (2020).

9. 9.

Minegishi, Y. *et al.* *Immunity* **25**, 745–755 (2006).

10. 10.

Fingerlin, T. E. *et al.* *Nature Genet.* **45**, 613–620 (2013).

11. 11.

World Health Organization. *Managing Epidemics: Key Facts About Major Deadly Diseases* (WHO, 2018).

| [Section menu](#) | [Main menu](#) |

- NEWS AND VIEWS
- 08 December 2021

# Benefits of megastudies for testing behavioural interventions

Trials of behavioural interventions are hard to compare, hampering policy decision-making. The effects of more than 50 interventions on exercise behaviour have been compared using an experimental design called a megastudy.

- [Heather Royer](#) 0

Over the past few decades, empirical work in the behavioural and social sciences has undergone a credibility revolution — a movement to use data to distinguish causality from correlation. Not surprisingly, one of the leading tools in this revolution is the randomized controlled trial (RCT). In these trials, participants are randomly allocated to groups that receive, or do not receive, an intervention. However, the conclusions that can be drawn from RCTs of behavioural interventions can be limited, and the results of different RCTs are difficult to compare. [Writing in Nature](#), Milkman *et al.*<sup>1</sup> show how a ‘megastudy’ — a collection of simultaneous RCTs — could overcome some of the limitations of individual RCTs. The authors’ megastudy tested the effectiveness of 54 behavioural interventions on gym attendance.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

## Rent or Buy

All prices are NET prices.

## **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

*Nature* **600**, 392–393 (2021)

doi: <https://doi.org/10.1038/d41586-021-03400-x>

## **References**

1. 1.

Milkman, K. L. *et al.* *Nature* **600**, 478–483 (2021).

2. 2.

Carrera, M., Royer, H., Stehr, M., Sydnor, J. & Taubinsky, D. *J. Health Econ.* **62**, 95–104 (2018).

3. 3.

Acland, D. & Levy, M. R. *Mgmt Sci.* **61**, 146–160 (2015).

4. 4.

Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R. & Wisdom, J. *Mgmt Sci.* **67**, 4139–4171 (2021).

5. 5.

Carrera, M., Royer, H., Stehr, M. & Sydnor, J. *Mgmt Sci.* **66**, 1890–1908 (2020).

6. 6.

Charness, G. & Gneezy, G. *Econometrica* **77**, 909–931 (2009).

7. 7.

Royer, H., Stehr, M. & Sydnor, J. *Am. Econ. J. Appl. Econ.* **7**, 51–84 (2015).

8. 8.

Andrews, I., Kitagawa, T. & McCloskey, A. *NBER Working Pap.* No. 25456 (Natl Bur. Econ. Res., 2019).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03400-x>

- Review Article
- [Published: 08 December 2021](#)

# The emergence, genomic diversity and global spread of SARS-CoV-2

- [Juan Li](#) ORCID: [orcid.org/0000-0003-0905-7188](https://orcid.org/0000-0003-0905-7188)<sup>1,2</sup> na1,
- [Shengjie Lai](#) ORCID: [orcid.org/0000-0001-9781-8148](https://orcid.org/0000-0001-9781-8148)<sup>3</sup> na1,
- [George F. Gao](#) ORCID: [orcid.org/0000-0002-3869-615X](https://orcid.org/0000-0002-3869-615X)<sup>4,5,6</sup> &
- [Weifeng Shi](#) ORCID: [orcid.org/0000-0002-8717-2942](https://orcid.org/0000-0002-8717-2942)<sup>1,2</sup>

*Nature* volume 600, pages 408–418 (2021)

- 31k Accesses
- 39 Altmetric
- [Metrics details](#)

## Subjects

- [SARS-CoV-2](#)
- [Viral infection](#)

## Abstract

Since the first cases of COVID-19 were documented in Wuhan, China in 2019, the world has witnessed a devastating global pandemic, with more than 238 million cases, nearly 5 million fatalities and the daily number of people infected increasing rapidly. Here we describe the currently available data on the emergence of the SARS-CoV-2 virus, the causative agent of COVID-19, outline the early viral spread in Wuhan and its transmission patterns in China and across the rest of the world, and highlight how genomic surveillance, together with other data such as those on human mobility, has helped to trace the spread and genetic variation of the virus and has also comprised a key element for the control of the pandemic. We pay particular attention to characterizing and describing the international spread of the major variants of concern of SARS-CoV-2 that were first identified in late 2020 and demonstrate that virus

evolution has entered a new phase. More broadly, we highlight our currently limited understanding of coronavirus diversity in nature, the rapid spread of the virus and its variants in such an increasingly connected world, the reduced protection of vaccines, and the urgent need for coordinated global surveillance using genomic techniques. In summary, we provide important information for the prevention and control of both the ongoing COVID-19 pandemic and any new diseases that will inevitably emerge in the human population in future generations.

[Download PDF](#)

## Main

On 31 December 2019, the Wuhan Municipal Health Commission reported an outbreak of pneumonia on its official website. Subsequently, scientists reported the discovery of a previously undescribed coronavirus obtained from samples of the respiratory system of some of these patients. This virus differed from all known coronaviruses including severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) and Middle East respiratory syndrome (MERS) coronavirus (MERS-CoV)<sup>1,2,3,4,5</sup>. The World Health Organization (WHO) named the disease coronavirus disease 2019 (COVID-19) and the International Committee on Taxonomy of Viruses named this new infectious agent SARS-CoV-2 (ref. <sup>6</sup>); the seventh coronavirus that can infect humans. SARS-CoV-2 rapidly spread through the globally, producing several variants of concern (VOCs) and developing into a major and devastating pandemic. Here we summarize our current understanding of the emergence, global spread and genetic diversity of SARS-CoV-2.

## The emergence of SARS-CoV-2

### SARS-CoV-2 related coronaviruses

Many of the early cases of COVID-19 in Wuhan, China, were associated with the Huanan Seafood Market<sup>2</sup>, which—because of the presence of wildlife at the market—was considered an obvious candidate for the location of the initial zoonotic (that is, cross-species transmission) event. However, none of the animals from the market (including rabbits, snakes, stray cats, badgers and bamboo rats) tested positive for SARS-CoV-2 (ref. <sup>7</sup>), and viral genome sequences of environmental samples from the market were not considered to occupy basal positions on the viral phylogeny (although the position of the rooting on the tree is uncertain)<sup>8</sup>. In addition, some of the early cases of COVID-19 in Wuhan were not epidemiologically linked to the market<sup>9</sup>, and some were linked to other markets<sup>10,11</sup>. Therefore, although it has not been resolved

fully, the current evidence suggests that the Huanan Seafood Market could be the location of an early ‘superspreading’ event.

From the earliest genomic comparisons, it was clear that SARS-CoV-2 had a genomic organization similar to SARS-CoV<sup>2</sup>. The spike proteins of both viruses have similar three-dimensional structures, suggesting that these viruses might use the same cell surface receptor—human angiotensin-converting enzyme 2(ACE2)<sup>2</sup>: this was soon confirmed *in vitro*<sup>4,12</sup> and using structural biology<sup>12,13</sup>. However, SARS-CoV-2 differs from SARS-CoV in two fundamental ways<sup>14</sup>. First, there are six amino acid positions in the receptor-binding domain (RBD) of the spike protein that mediate the attachment of the SARS-CoV and SARS-CoV-2 spike proteins to the human ACE2 receptor<sup>15</sup>. However, amino acids at five of the six positions differed between SARS-CoV and SARS-CoV-2 (refs. <sup>2,14</sup>). Notably, such differences caused SARS-CoV-2 to have a higher binding avidity to the human ACE2 receptor<sup>11</sup>, and may have contributed to the higher transmissibility of SARS-CoV-2 compared with SARS-CoV. Second, there is a 12-nucleotide (nt) insertion at the cleavage site of the spike protein of SARS-CoV-2 that has not yet been identified in closely related betacoronaviruses, but that has a complex evolutionary history across the coronaviruses as a whole, indicating that it is evolutionarily volatile<sup>16</sup>. This insertion encodes four amino acids—PRRA—that can be recognized by the protease furin, which is extensively expressed in different tissues and organs<sup>17</sup>. This insertion may decrease the overall stability of the SARS-CoV-2 spike, thereby facilitating the adoption of the open conformation that is required for the binding of the spike to human ACE2 (ref. <sup>18</sup>); SARS-CoV-2 without this furin-cleavage site shows reduced replication in a human respiratory cell line and was attenuated in laboratory animals<sup>19</sup>. Notably, amino acid substitutions have been documented at all four positions in the PRRA motif, with a P-to-H substitution (HRRA) identified in more than 487,000 viral genomes as of June 2021.

SARS-CoV-2—like many other members of the genus *Betacoronavirus* (including SARS-CoV) in the *Coronaviridae* family—seemingly has its evolutionary roots in those viruses that commonly infect bats<sup>2</sup>. Not surprisingly, shortly after the identification of SARS-CoV-2, a close relative of SARS-CoV-2 was described; RaTG13 was identified from a bat (*Rhinolophus affinis*) sample obtained in Yunnan Province, China, in 2013 (ref. <sup>4</sup>). Notably, this sample was collected from a mine cave to which four workers were sent to clean bat faeces and who subsequently developed severe pneumonia<sup>20</sup>. Although RaTG13 exhibits 96.2% sequence identity to SARS-CoV-2 at the scale of the whole genome, it does not possess similar RBD or cleavage-site sequences. Further analyses suggest that RaTG13—rather than SARS-CoV-2—was a recombinant virus, and the two virus lineages probably diverged more than 30 years ago<sup>21</sup>. Therefore, the SARS-CoV-2 RBD was an ancestral trait shared with bat viruses<sup>21</sup>.

Subsequently, a number of groups reported the identification of SARS-CoV-2-related coronaviruses in Malayan pangolins (*Manis javanica*), which were smuggled to Guangxi and Guangdong provinces, China<sup>22,23</sup>. These pangolin coronavirus genomes exhibited 85.5–92.4% sequence similarity to SARS-CoV-2 (ref. <sup>22</sup>). Notably, however, these pangolin-derived coronaviruses formed two sublineages, with the Guangdong sublineage clustering with RaTG13 and SARS-CoV-2 and sharing 97.4% amino-acid similarity to SARS-CoV-2 in the RBD, with identical amino acids at the five critical residues of the RBD. Furthermore, the Guangdong pangolins appeared to have a similar disease manifestation to people with COVID-19 (ref. <sup>24</sup>). Therefore, although the role—if any—of pangolins in the origin of SARS-CoV-2 and the ecology of coronaviruses in general is unknown, it is clear that coronaviruses exist in wildlife and that these viruses possess SARS-CoV-2-like RBDs and have a high binding avidity to hACE2.

Furthermore, a previously undescribed bat coronavirus—RmYN02—was reported, which had been collected during routine surveillance of *Rhinolophus malayanus* bats in Yunnan Province on 25 June 2019 (ref. <sup>25</sup>). RmYN02 shared 97.2% sequence identity with SARS-CoV-2 in open-reading frame (ORF) 1ab. ORF1ab is the largest in coronaviruses with a length of approximately 21,300 nt. In June 2021, we reported four SARS-CoV-2-related coronaviruses genomes from Yunnan Province<sup>26</sup>. Of these, RpYN06, found in *Rhinolophus pusillus*, exhibited 94.5% sequence identity to SARS-CoV-2. However, the genome—excluding the spike gene, which has a history of recombination—had a similarity to SARS-CoV-2 of 97.2%, making it the closest related genomic backbone to SARS-CoV-2 identified to date. The other three SARS-CoV-2-related coronaviruses were more distantly related to SARS-CoV-2. However, they carried a genetically distinct spike genes encoding proteins that could bind to the human ACE2 receptor in vitro, albeit weakly.

SARS-CoV-2-like coronaviruses have also been identified in bat populations from other parts of Asia, including Japan<sup>27</sup>, Cambodia<sup>28</sup> and Thailand<sup>29</sup>. Notably, although two betacoronaviruses (STT182 and STT200) from *Rhinolophus shameli* bats sampled in 2010 from Cambodia shared 92.6% nucleotide identity with SARS-CoV-2 across the genome as a whole, they share five of the six critical RBD sites observed in SARS-CoV-2 and the Guangdong pangolin coronavirus<sup>28</sup>. In September 2021, a preprint described a number of SARS-CoV-2-related coronaviruses identified in Laos, including BANAL-52 from *R. malayanus*, BANAL-103 from *R. pusillus* and BANAL-236 from *Rhinolophus marshalli*, which only possessed one or two amino acid mismatches at the seventeen residues that interact with human ACE2<sup>29</sup>. In particular, the RBDs of these viruses could bind as efficiently to the human ACE2 protein as could the SARS-CoV-2 Wuhan strain from the early stage of the pandemic.

## Emergence pathways of SARS-CoV-2

There are several hypotheses regarding the origin and emergence of SARS-CoV-2 that have been thoroughly clarified in the WHO–China joint report<sup>7</sup>. These contradictory hypotheses have raised standing debates, with the central point being two competing hypotheses: zoonotic emergence (including direct zoonotic introduction or introduction through an intermediate host) and a laboratory escape. The discovery of more and more SARS-CoV-2-related coronaviruses from wild animals provides evidence for a zoonotic origin of SARS-CoV-2 (refs. [4](#),[22](#),[23](#),[25](#),[26](#),[27](#),[28](#),[29](#),[30](#)). Notably, all of the SARS-CoV-2-related coronaviruses mentioned above are evidently not the direct ancestor of SARS-CoV-2. Any such direct ancestral virus—which has yet to be identified—would be expected to exhibit more than 99% similarity to SARS-CoV-2 across the genome as a whole. However, the discovery of these viruses again highlights that more-closely related viruses in bats and other wildlife species will be identified with enhanced sampling in a broader geographical region, including most parts of Southeast Asia, which has a high diversity of *Rhinolophus* species<sup>26</sup>. As it has seldomly been found that a bat coronavirus is able to efficiently transmit among humans without adaptation and repeated human–animal contacts<sup>10</sup>, introduction through an intermediate host, such as raccoon dogs, is more likely than a direct zoonotic introduction.

Whether SARS-CoV-2 was introduced through a laboratory accident or whether it has been genetically manipulated is highly debatable. After a thorough analysis of the genetic characterizations of SARS-CoV-2 from both the early and later stages of the pandemic, as well as its close relatives from wild animals, many researchers in the global scientific community have reached the consensus that SARS-CoV-2 is unlikely to have escaped a laboratory and there is no scientific evidence that SARS-CoV-2 has been genetically manipulated<sup>10</sup>. However, the exact spillover event and emergence process of SARS-CoV-2 is still unclear, and more information from the earliest stage of the epidemic is clearly important to understand how SARS-CoV-2 came into contact with people.

## Global genetic diversity of SARS-CoV-2

### Genomic surveillance of SARS-CoV-2

Mutations are a natural part of the replication cycle of any RNA virus, leading to the diversification of viral lineages when coupled with inter-host transmission. This is also true for SARS-CoV-2, even though coronaviruses contain certain proofreading mechanisms that enhance genome fidelity<sup>31</sup>. Genomic surveillance has generated an unprecedented amount of sequencing data for a single virus (Box 1), and has proven an essential tool<sup>32,33</sup> for tracing the spread of SARS-CoV-2 at various scales, from individual transmission events to the intercontinental spread of the virus. In addition, it

has had a central role in monitoring the evolution of SARS-CoV-2 and identifying new variants with enhanced transmissibility and/or pathogenicity, decreased susceptibility to therapeutic agents and that are capable of evading natural or vaccine-induced immunity (Fig. 1). Genomic surveillance has demonstrated the effectiveness of tracking local transmission events, recognizing importation sources and superspreading events in Australia<sup>34,35</sup>, for informing public-health decision-making in the Netherlands<sup>36</sup>, and for adopting social-distancing measures to reduce viral spread in Israel<sup>37</sup>. In January 2021, du Plessis and colleagues described the analysis of 50,887 SARS-CoV-2 genomes<sup>38</sup>, quantifying the viral genetic structure of the UK epidemic at a fine scale, including the size, spatiotemporal origins and persistence of lineages as well as the effect of intervention measures.

**Fig. 1: Phylogenetic tree of SARS-CoV-2 lineages globally and the temporal distribution of major sequence variants.**

---

 **figure1**

The phylogenetic analysis was performed using full-length genome sequences of SARS-CoV-2 collected from GISAID as of 12 May 2021. A maximum likelihood tree of 1,715 representative high-quality SARS-CoV-2 sequences carrying specific accumulative mutations was estimated using RAxML<sup>159</sup>, with 1,000 bootstrap

---

replicates and the GTR nucleotide substitution model. The major VOCs (Alpha to Delta) are shown in orange, and the major variants of interests (Epsilon to Lambda) are shown in purple. Both the thickness of each branch in the phylogenetic tree and the shading from light to dark in the heat map indicate the number of sequences carrying specific sets of mutations. Specific nucleotide substitutions are highlighted on the major branches of the tree. The branches with the D614G substitution are coloured blue.

Below, we use Guangdong Province, China and the USA as examples to illustrate how genomic surveillance has facilitated our understanding of this pandemic.

## Guangdong, China

Guangdong is a populous province in Southeast China, with a resident population of more than 100 million people. After the SARS-CoV outbreak, believed to have originated in Guangdong<sup>39</sup>, long-term reforms in public-health agencies have greatly improved the infrastructures and enhanced the capacity of disease control and prevention. The first case of COVID-19 in Guangdong had an onset of symptom on 1 January and was reported on 19 January 2020 (refs. [11,40](#)). Like many other Chinese provinces, Guangdong experienced three phases—domestic importation, local community transmission and international importation—with an epidemic peak in early February 2020 (ref. [40](#)). Large-scale surveillance (around 1.6 million tests by 19 March 2020 identifying 1,388 cases of COVID-19) and intervention measures were implemented from the beginning of the outbreak, and after 22 February 2020 no more than one case a day was reported<sup>40</sup>. The genomic epidemiology of SARS-CoV-2 in Guangdong showed that most of the infections before March were imported from Hubei Province, and, in particular, Wuhan. Although some early cases were caused by community transmission, local transmission chains were limited both in size and duration<sup>40</sup>. These results highlight the efficacy of intensive testing and contact tracing even in such a densely populated urban region. Intensive surveillance also identified two SARS-CoV-2 variants with deletions in the spike gene<sup>41</sup>. In addition, the Guangdong Centers for Disease Control and Prevention (CDC) successfully identified the imported Alpha and Beta variants on 2 January 2021 (ref. [42](#)) and 6 January 2021 (ref. [43](#)), respectively.

## The USA

The first case of COVID-19 in the USA (sequence WA1) was reported on 20 January 2020—a traveller from Wuhan<sup>44</sup>. By 15 February 2020, the number of laboratory-confirmed and clinically diagnosed cases of COVID-19 had reached 15 (ref. [45](#)). By combining multiple sources of information, Worobey and colleagues showed that

transmission of the WA1 (belonging to lineage A) lineage was successfully contained, and the subsequent larger outbreaks in Washington state might have been caused by multiple independent introductions of the virus from China in late January or early February 2020 (ref. <sup>46</sup>). However, evidence from various studies revealed that the early viruses that were present between 29 February and 18 March 2020 in New York City were imported from Europe and other parts of the USA by multiple, independent introductions<sup>47</sup>. In addition, cryptic transmission and a prolonged period of unrecognized community spread has been documented in northern California<sup>48</sup>, Washington state<sup>49</sup> and New York City<sup>50</sup> from late January to March 2020. For example, SARS-CoV-2 sequences sampled from Connecticut during 6–14 March 2020 group with those from Washington state, highlighting long-distance domestic transmission<sup>51</sup>. Genomic surveillance in Dane and Milwaukee counties in Wisconsin between March and April 2020 provided evidence for reduced viral spread after a state-wide ‘safer at home’ order<sup>52</sup>. Together, these genomic surveillance studies clearly illustrate the early transmission of SARS-CoV-2 and highlight the efficacy of intensive testing, contact tracing and decreasing public gatherings in containing the spread of SARS-CoV-2.

## **Box 1 Sources of SARS-CoV-2 genomic data and surveillance**

### **The GISAID database**

There are more than 2.8 million complete SARS-CoV-2 genomes and metadata available from the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV as of August 2021 (<https://www.gisaid.org/>). Useful tools, including BLAST search, phylogenetic trees, PrimerChecker, spike glycoprotein mutations and surveillance of emerging variants are provided, and related analyses are constantly updated.

### **The NCBI database**

More than 1.1 million SARS-CoV-2 nucleotide records and 900,000 SRA runs have been deposited in the National Center for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>) databases. The NCBI SARS-CoV-2 Resources (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) also provide comprehensive access to other related data sources and numerous online analysis tools.

### **The CNBC/NGDC database**

The Chinese National Bioinformatics Center/National Genomics Data Center (CNBC/NGDC; <https://bigd.big.ac.cn/ncov/>) database integrates the SARS-CoV-2 genomes and related metadata from other sources, such as GISAID, NCBI, GWH

(Genome Warehouse, <https://bigd.big.ac.cn/gwh/>), NMDC (National Microbiology Data Center) and CNGB (China National GeneBank)<sup>161</sup>. It provides a variety of useful online analysis tools, including sequence integrity and quality assessments, spatiotemporal dynamics, haplotype network, variant distribution, molecular mutations and published clinical trials.

### Pango lineages

Pango lineages (<https://cov-lineages.org/>) is a useful nomenclature system for SARS-CoV-2 genomes. As of August 2021, the Pango system contains more than 1,500 designated lineages covering all of the SARS-CoV-2 sequences from GISAID. Web-based or open-source code of applications such as Pangolin, Scorpio, Civet, Polecat are internally developed to identify clusters. Using the Pangolin web interface (<https://pangolin.cog-uk.io/>), sequences uploaded by users can be assigned to the most likely lineage based on the Pango dynamic nomenclature<sup>162</sup>. Information about the SARS-CoV-2 variants is also provided.

### Nextstrain SARS-CoV-2 resources

Genomic epidemiological analysis of global SARS-CoV-2 is continually updated on the open-source platform Nextstrain (<https://nextstrain.org/sars-cov-2/>), based on the genomic data from GISAID. It provides a variety of visualization options for users. The nucleotide and amino acid diversity of the spike gene and protein, and the frequencies of the Nextstrain clades are provided and updated. In addition, Nextclade can perform clade assignment, mutation calling and sequence quality checks for the SARS-CoV-2 sequences uploaded by users.

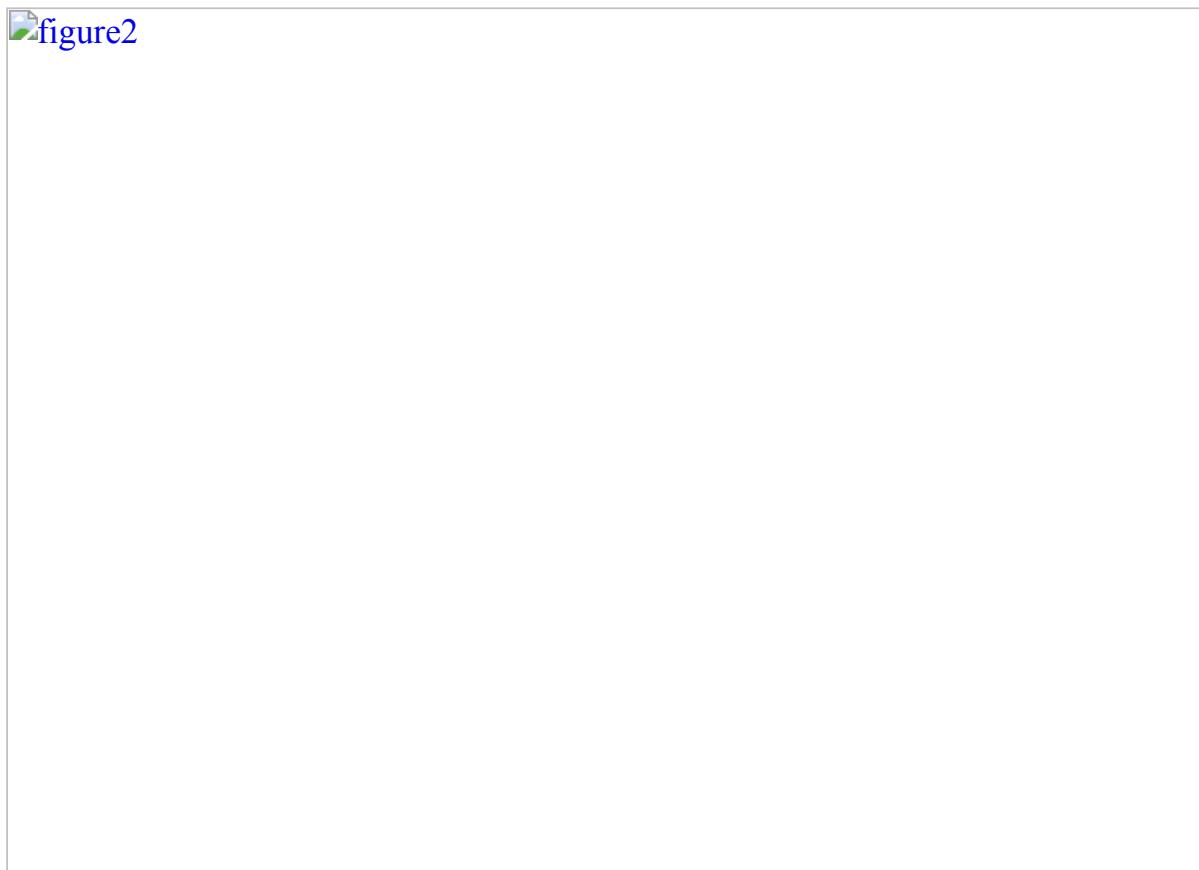
### Mutational diversity of SARS-CoV-2

By January 2021, approximately 25,000 out of the 29,800 sites (the length of the complete SARS-CoV-2 genome) have been shown to carry mutational differences (<https://bigd.big.ac.cn/ncov/>), and it has been estimated that approximately two mutations are fixed in the SARS-CoV-2 genome per month<sup>46,53,54</sup>. Although most of these mutations represent standard replication errors, host-dependent RNA editing may also shape the short- and long-term evolution of SARS-CoV-2. Indeed, the SARS-CoV-2 genome is characterized by frequent biased C-to-U hypermutation that is probably due to a human APOBEC-like editing process<sup>55,56</sup>.

Similar to other coronaviruses, the spike protein of SARS-CoV-2 contains important antigen epitopes<sup>57,58</sup>. As such, mutations in the spike protein will probably affect the receptor-binding efficiency, potentially lead to immune escape and may even weaken vaccine efficacy. The first notable mutation was A23403G, which caused the D614G

amino acid substitution in the spike protein. This mutation might have arisen separately as early as late January 2020 in China and later in Europe, representing an interesting evolution of a mutation of convergence, and the frequency of this mutation greatly increased during the outbreak in Europe<sup>59,60</sup>. There is now compelling evidence that D614G has increased virus infectivity and transmissibility<sup>59,60,61,62,63,64</sup>, and molecular epidemiological studies suggest that this mutation increased the basic reproduction number ( $R_0$ ) from 3.1 (614D) to 4.0 (614G)<sup>60</sup>. In addition, a so-called ‘cluster V’ (also called B.1.1.298) SARS-CoV-2 variant was identified in Danish mink that also carried mutations in the spike protein, including Y453F, I692V, M1229I and the deletion of two amino acids ( $\Delta$ 69– $\Delta$ 70)<sup>16,65</sup> (Fig. 2).

**Fig. 2: SARS-CoV-2 spike mutations in the Alpha, Beta, Gamma, Delta and mink cluster V variants.**



Three-dimensional structures were modelled with the Swiss-Model program using the spike protein of SARS-CoV-2 (PDB: 7CWU.1.G) as a template. Left, red spheres represent the mutations found in the Alpha<sup>160</sup>, Beta<sup>69</sup>, Gamma<sup>146</sup> and Delta<sup>146</sup> VOCs, as well as the mink cluster V variants<sup>65</sup>. The amino acid positions of all of the strains are numbered according to the template. Right, the surfaces of the six amino acid residues (L455, F486, Q493, S494, N501 and Y505) at the RBD are coloured cyan.

The molecular surfaces of the mutations in the Alpha (purple), Beta (blue), Gamma (yellow), Delta (green), and mink cluster V (pink) variants are highlighted. \*Not all Alpha variants have the E484K and S494P mutations. #Not all Delta variants possess the G142D mutation. It should be noted that we only use this figure to highlight the locations of the mutations in the variants based on the three-dimensional structure of one ancestral Wuhan strain (NC\_045512), and this figure does not necessarily represent the true three-dimensional structure of the variants.

Not surprisingly, as the number of cases of COVID-19 continued to rise, mutational variants with a likely greater effect on fitness have also emerged, including some that might result in immune escape. Indeed, there are putative escape mutations to the ten human monoclonal antibodies that target the SARS-CoV-2 RBD<sup>66</sup>. Of particular note are the major SARS-CoV-2 VOCs that arose in late 2020: Alpha (also known as B.1.1.7 and VOC-202012/01), Beta (also known as B.1.351 and 501Y.V2), Gamma (also known as P.1) and Delta (also known as B.1.617.2); these VOCs were first identified in the UK<sup>67,68</sup>, South Africa<sup>69</sup>, Brazil<sup>70,71</sup> and India<sup>72</sup>, respectively (Box 2 and Figs. 1, 2).

The emergence of these variant lineages has raised concerns that the virus has entered a new phase in its evolution<sup>73,74,75</sup>, characterized by ongoing immune escape in the face of increasing levels of infected hosts that probably affects vaccine efficacy, as well as the possibility of selection for increased transmission due to the imposition of nonpharmaceutical interventions (NPIs)<sup>74</sup>. The Alpha variant has been associated with increased rates of virus population growth<sup>67,68</sup> and has been reported to be able to escape neutralization by most monoclonal antibodies targeting the N-terminal domain (NTD) of the spike protein<sup>76</sup>. However, there is no widespread escape of the Alpha variant from monoclonal antibodies or antibody responses generated by natural infection or vaccination<sup>76,77,78</sup>, such that its spread may instead reflect increased transmissibility. In particular, some of the Alpha variants acquired additional mutations in the spike protein, especially E484K, and exhibited a substantial loss of sensitivity to the neutralizing activity of vaccine-elicited antibodies and resistance to neutralization by monoclonal antibodies in COVID-19-convalescent plasma<sup>79</sup>. More worryingly, the Beta variant can escape neutralization by most RBD-targeting monoclonal antibodies and substantially escape from neutralizing antibodies from COVID-19-convalescent plasma<sup>76,80,81</sup>. Similarly, the Gamma variant shows marked decreases in neutralization with post-vaccination sera<sup>82</sup>; although, surprisingly, it is considerably less resistant to naturally acquired or vaccine-induced antibody responses than the Beta lineage<sup>83</sup>. Furthermore, neutralization of the Delta lineage is reduced when compared with ancestral circulating strains<sup>77,78</sup>, and convalescent sera from patients infected with the Beta and Gamma variants show a markedly higher reduction in neutralization of the Delta lineage<sup>77</sup>.

In addition to nucleotide substitutions, the SARS-CoV-2 genome has experienced many deletion events. For example, some viruses from Singapore and Taiwan, China carried a 382-nt deletion truncating ORF7b and covering almost the entire ORF8 sequence<sup>84,85,86</sup>. This variant showed considerably higher replicative fitness in vitro than the wild-type virus<sup>84</sup>, but seemed to be associated with a milder infection clinically<sup>85</sup> and has not been reported in 2021. Su and colleagues described other ORF7b/8 deletions of various lengths, including viruses from Australia (138 nt), Bangladesh (345 nt) and Spain (62 nt)<sup>84</sup>. Long deletion events were also found in clinical samples from Beijing, with a 120-nt deletion in ORF7a and a 154-nt deletion in ORF8 (ref. <sup>87</sup>).

## Box 2 Genetic characterizations of the major VOCs

### The Alpha variant

The Alpha variant is defined by 17 amino-acid-altering mutations (14 non-synonymous mutations and 3 deletions), including 8 in the spike protein (Figs. [1](#), [2](#)). Notably, three of these mutations are of potential biological importance: N501Y, P681H and the deletion of two amino acids 69 and 70 ( $\Delta 69-\Delta 70$ )<sup>67,68</sup>. Notably, this new variant has increased infectiousness across all age groups, and is 43% to 90% more transmissible than previously circulating strains<sup>67,68</sup>. In addition, infection with the Alpha variant has the potential to cause substantial additional mortality, with an increased risk of death of 32–104% (ref. [163](#)). However, there are also reports of no association between this variant and increased severity<sup>164,165</sup>. As of 10 August 2021, 185 countries, territories and areas have identified this variant<sup>166</sup> (Fig. [3b](#) and Extended Data Fig. [1a](#)).

### The Beta variant

The Beta variant is characterized by eight lineage-specific mutations in the spike protein, including three at important residues in the RBD (K417N, E484K and N501Y)<sup>69</sup> (Figs. [1](#), [2](#)). In addition to South Africa, 135 additional countries, territories and areas have also reported the identification of this variant as of 10 August 2021 (Fig. [3b](#) and Extended Data Fig. [1b](#)), with community transmission mainly found in Africa, Europe and North America<sup>166</sup>.

### The Gamma variant

The Gamma variant contains a number of potentially important mutations, such as K417T, E484K, and N501Y in the spike protein<sup>70,71</sup> (Figs. [1](#), [2](#)). The Gamma variant might be 1.7–2.4-fold more transmissible than previous (non-Gamma) strains in Brazil. As of 10 August 2021, identification of this variant has been reported in 81

countries, territories and areas (Fig. 3b and Extended Data Fig. 1c), most of which are located in the Americas and Europe<sup>166</sup>.

## The Delta variant

The Delta variant contains several important amino acid mutations in the spike protein, including three-amino acid-altering mutations (two deletions at positions 156 and 157 ( $\Delta$ 156– $\Delta$ 157)), one substitution (R158G) in the NTD, L452R, T478K, and P681R<sup>167</sup> (Figs. 1, 2). The Delta variant itself has shown ongoing evolution and a so-called ‘Delta plus’ variant with an additional K417N substitution in the spike protein was identified in India in June 2021 (refs. 138,168).

Despite their independent emergence (Fig. 1), the Alpha, Beta, and Gamma variants have the N501Y mutation found in the mouse-adapted SARS-CoV-2 variant<sup>169</sup>. In addition, the Beta and Gamma lineages share E484K<sup>67,68,70,71</sup>, which was also identified in the late, rather than early, Alpha variants<sup>79</sup>.

# Global spread of SARS-CoV-2

## Initial spread of SARS-CoV-2 in China

Generally, China experienced three distinct phases of SARS-CoV-2 transmission: (1) the initial rapid spread in Wuhan; (2) seeding from Wuhan to cause community transmission in other regions of China; and (3) sporadic outbreaks caused by international importations after China controlled the first wave<sup>40,87</sup>.

## Early spread of SARS-CoV-2 in Wuhan

The initial SARS-CoV-2 outbreak in Wuhan can itself be divided into three phases<sup>88</sup>: (1) rapid transmission before the implementation of the large-scale population ‘lockdown’ of the city on 23 January 2020 (ref. 9), with an estimated effective reproduction number ( $R_e$ ) of 3.5 (95% credible interval, 3.4–3.7) during this period<sup>89</sup>; (2) reduction of the rate of virus transmission during the period 23 January–1 February 2020 (through lockdown and home quarantine), producing an average  $R_e$  of 1.2 (95% credible interval, 1.1–1.3)<sup>89</sup>; and (3) the interruption of transmission through intensified stringent interventions during 2–16 February 2020 (centralized isolation and treatment of cases of COVID-19) and 17 February–8 March 2020 (community screening). Population-based serological surveys conducted during March–May 2020 revealed that the overall seropositivity rate in Wuhan was 3.2–4.4% (refs. 90,91,92,93), indicating that many cases went undetected due to asymptomatic and mild infections

and the limited laboratory-diagnosis capacity during the early stages of the outbreak<sup>89,94,95</sup>. However, city-wide nucleic acid screening of SARS-CoV-2 between 14 May and 1 June 2020 among nearly 10 million residents of Wuhan only found around 300 individuals who had asymptomatic infections after the lockdown was lifted on 8 April 2020 (ref. 96) and no symptomatic local cases related to the initial wave have been reported in the city after 10 May 2020.

## Spread from Wuhan to other provinces

The coincidence of the emergence of SARS-CoV-2 and the large-scale seasonal migration (*Chunyun*, starting from 10 January 2020) for the Chinese Lunar New Year holiday probably exacerbated the seeding of the virus across China<sup>97,98</sup>. Movement restrictions from Wuhan, the key transportation hub in central China, commenced on 23 January 2020, and reduced the peak population numbers leaving the city 2 days before the Lunar New Year. Unfortunately, however, the disease had spread to every province in mainland China by this time<sup>99,100</sup>. In general, after the rapid implementation of stringent and integrated NPIs, the  $R_e$  in provinces outside Hubei decreased below the epidemic threshold (1.0) from 8 February 2020 (ref. 101). Compared with Wuhan, the seropositivity rate in cities outside Wuhan was much lower. According to a national COVID-19 sero-epidemiological survey in China during March–May 2020 (ref. 92), only 0.44% of the sampled population in other cities of Hubei were positive, and only 2 out of more than 12,000 people outside Hubei tested positive, suggesting that SARS-CoV-2 transmission was well contained across the country during the first wave<sup>99,102,103</sup>.

## Frequent international importation events

More than 6,000 incoming travellers from abroad who were infected with SARS-CoV-2 had been reported in mainland China by 15 June 2021, although reverse-transcriptase–polymerase-chain-reaction (RT–PCR) testing at the border control and a 14-day centralized quarantine implemented in China since March 2020 greatly reduced any transmission risk. For example, in Guangzhou, Guangdong Province in southern China, 73.5% of the imported positive cases were detected at the immigration checkpoint and 19.0% during centralized quarantine in hotels<sup>104</sup>. Although SARS-CoV-2 is predominantly associated with respiratory transmission, since June 2020, multiple Chinese provinces have detected SARS-CoV-2 RNA or live virus on packages of frozen products<sup>105</sup>. Indeed, cold-chain food or package contamination was proposed to have triggered the resurgence in Beijing in June 2020 (ref. 106) as well as other sporadic outbreaks in China<sup>105</sup>, although this warrants further investigation. It is notable that the number of confirmed cases was low in the Xinfadi outbreak, Beijing,

in June 2020. Similarly, all of the COVID-19 outbreaks in China triggered by international inbound travellers were small-scale, with a few sustained cases. This was mainly due to the citywide, grid-based mass-screening protocol using RT-PCR testing<sup>107</sup>.

## **Intercontinental spread of SARS-CoV-2**

## From China to other regions

The global spread of SARS-CoV-2 shows how rapidly geographically disparate countries can be reached by an emerging pathogen<sup>108,109</sup> (Fig. 3a). Two distinct transmission phases of international exportations of SARS-CoV-2 were identified at the early stage of the pandemic<sup>110</sup>. In the first phase, many international airline passengers left Wuhan for hundreds of destinations across the world during the two weeks before the Wuhan lockdown<sup>92</sup>. Cities across Asia, Europe and North America were the main destinations and reported several imported cases during the early stage of the COVID-19 outbreak<sup>109,111</sup>, and the WHO declared a Public Health Emergency of International Concern on 30 January 2020. Containment of the outbreak in China and, in particular, the implementation of travel restrictions since late January 2020 considerably reduced the further spread of SARS-CoV-2 outside China<sup>99,100,102,112,113</sup>.

**Fig. 3: Global spread of SARS-CoV-2 and cases reported across countries.**



**a**, The date of the first reported case of COVID-19 in each country, territory or area, though the origin of SARS-CoV-2 has not been determined for almost two years. The areas without data are shown in grey. **b**, Reports of VOCs (now denoted VOC Alpha to Delta) based on records published in the COVID-19 weekly epidemiological update by the WHO (<https://covid19.who.int/>), as of 10 August 2021. **c**, The seven-day rolling average of the number of confirmed cases of COVID-19 reported by continent. The orange vertical dashed line indicates the date of COVID-19 declared as a pandemic by

the WHO. **d**, **e**, The weekly proportion of case number in the top-50 ranked countries with the highest number of cases of COVID-19 (**d**) and the available mobility data (**e**), as of 8 August 2021. The weekly proportion was calculated as the case count in a specific week and country, divided by the total number of cases reported in each country. **e**, The changes in human mobility (by 8 August 2021) in the 50 countries presented in **d**, compared to the normal mobility from 3 January to 6 February 2020. Each row in **d** and **e** represents a country, grouped by continent and then sorted by the latitudes of capital cities from north to south (the country list is available in Supplementary Table 1). The grey dotted vertical lines in **d** and **e** from left to right indicate the first week of April, July and October in 2020, and January, April and July in 2021, respectively. The dataset of case numbers was obtained from the data repository collated by the Johns Hopkins University ([github.com/CSSEGISandData/COVID-19](https://github.com/CSSEGISandData/COVID-19)). The anonymized and aggregated data of population mobility in transit stations were obtained from the Google COVID-19 Community Mobility Reports ([www.google.com/covid19/mobility/](https://www.google.com/covid19/mobility/)). The administrative boundary maps were obtained from Natural Earth ([www.naturalearthdata.com](https://www.naturalearthdata.com)).

## From Europe to other regions

However, international travel outside China from mid-February to late-March 2020 facilitated the second phase of international SARS-CoV-2 spread and onward transmissions<sup>110,114</sup>, with the epicentre quickly shifting to the Middle East<sup>115</sup> and Europe (Fig. 3c). Although France was the first country to identify cases of COVID-19 in Europe, Italy soon became the first major hotspot in the continent<sup>111,112,116,117</sup>, whereas Spain, Belgium and the UK reported the highest numbers of deaths in Europe during the first wave<sup>118</sup>. The virus exported from Europe acted as a major source of global spread<sup>47</sup>, and the WHO eventually declared a pandemic on 11 March 2020. Countries quickly placed restrictions on flights from Europe during March–April 2020, although these measures could not fully prevent introduced transmission<sup>75,114</sup>.

By late March 2020, cases surged in the USA, with North America becoming the global epicentre<sup>119,120</sup>. By the end of 2020, the total number of confirmed cases recorded in the USA had passed 20 million, including more than 350,000 reported deaths. Although the first case of COVID-19 in the USA was reported in a traveller returning from China on 20 January 2020 (ref. 44), phylogenetic evidence suggests that importations from Europe mainly contributed to the wide spread of the virus across the country<sup>110,119</sup>. Latin America and south Asia have also been badly affected. SARS-CoV-2 was confirmed in Brazil on 25 February 2020 and a month later it was found in every state, with confirmed cases exceeding 1 million on 19 June 2020 (refs. 121,122). Although the first case of COVID-19 was confirmed in India on 30 January 2020 and

the situation was seemingly under control until the end of March 2020 (ref. [123](#)), India has reported the second highest number of cases of COVID-19 since September 2020 (ref. [124](#)). Most African countries experienced community transmission by 31 May 2020, with most imported cases returning from Europe and the USA[125](#), and it is believed that the disease is generally underreported across Africa due to the limited testing and healthcare capacities[126,127,128,129](#).

## Spread of secondary waves across countries

NPIs—such as travel restrictions, case isolation and contact tracing, physical distancing, face covering, hand washing and even the closures of businesses and schools—have been widely implemented to reduce the transmission of SARS-CoV-2 (refs. [112,130,131](#)). Full or partial lockdowns during specific periods have also been imposed in many countries[118](#). Although the effectiveness of different interventions and their combinations have varied, these measures have had an important role in the response to the first wave of the pandemic[132,133](#).

Unfortunately, after the relaxation of these interventions, an increase in population movements and the spread of new variants with a higher transmissibility, a new wave of infections has swept through many nations since October 2020 (refs. [134,135,136](#)) Fig. [3d,e](#) and Supplementary Table [1](#)). The first US wave in 2020 mainly affected the northeast of the USA[137](#), whereas the second wave in summer 2020 mainly hit the south and west, and almost every state has seen a spike in cases during the third wave since October 2020 (ref. [138](#)). Brazil has experienced a major second wave since November 2020 and even had a death toll second only to the USA in early 2021 (ref. [139](#)). New SARS-CoV-2 variants also spread throughout Europe after travel resumed in summer 2020[140,141](#), with the highest daily number of cases recorded in many countries between October 2020 and March 2021. After NPIs were implemented together with a second or third lockdown, and combined with ongoing and large-scale vaccination efforts, many countries passed the second wave by the end of May 2021. This has reduced the pressure on healthcare systems and given countries time to vaccinate people at the greatest risk of severe disease[142](#).

However, the emergence and rapid spread of various SARS-CoV-2 VOCs and variants of interest (VOIs) that are more contagious and/or potentially evade immunity has triggered new waves in many countries (Fig. [3b](#) and Extended Data Fig. [1](#)). For example, India has experienced a major second wave from March to June 2021, mostly due to the Delta variant. As of 10 August 2021, a total of 142 countries, territories and areas across the world have reported the Delta variant[72](#) (Extended Data Fig. [1d](#)), including countries with mass vaccination of their populations, such as the UK and Israel[143](#). In particular, community transmission of this variant has been

reported in many countries<sup>72</sup>. In mid-June 2021, the WHO declared that the Delta variant has displaced most of the other VOCs and has become the dominant lineage across the world<sup>143,144</sup>.

## Challenges and outlook

Even though it is of vital importance to the prevention of future emerging infectious diseases that will inevitably affect human populations, our current understanding of the initial SARS-CoV-2 spillover event is limited. Although the closest relatives to SARS-CoV-2 are found in horseshoe bats, it remains unclear whether the virus directly moved from bats to humans or was passed through an intermediate animal host—as was the case for previous coronavirus epidemics—although the latter seems more reasonable<sup>7,10</sup>.

The genomic surveillance of SARS-CoV-2 is by far the largest pathogen genomic sequencing project undertaken, with more than 2.8 million complete genomes generated as of August 2021. This endeavour has played an essential part in the prevention and control of COVID-19 and shed light on the transmission patterns of SARS-CoV-2 at different scales, such as the time and source of the introduction events, the spatiotemporal characterizations of local spread, the role of superspreading events, and the viral factors associated with the fitness, transmissibility, infectivity and disease severity. Of particular note is the identification of the major SARS-CoV-2 VOCs, as well as several variants of interests (denoted Epsilon to Lambda)<sup>145,146</sup> that emerged in different countries and have caused an increased proportion of cases both locally and globally.

The emergence of these SARS-CoV-2 variants has shaped the complex global transmission dynamics of COVID-19. More importantly, there is mounting evidence<sup>147</sup> that these SARS-CoV-2 variants are able to cause decreases in neutralizing titres from patients who recovered from COVID-19 and vaccine recipients, and escape neutralization by the monoclonal antibodies that target the NTD and RBD of the spike protein to various degrees. However, genomic surveillance would be more informative if coupled with a system for the risk assessment and phenotyping of these mutations. For example, the infectivity and antigenicity of 106 mutations in the SARS-CoV-2 spike was assessed using pseudotyped viruses<sup>148</sup>. Deep mutational scanning has also been used to assess all single amino acid variants of the SARS-CoV-2 spike protein<sup>66,149</sup>. In addition, more and more data on antigenic variations of the SARS-CoV-2 variants, with different sets of single amino acid mutations, to monoclonal antibodies and vaccines are available. A risk assessment system that integrates pathogen surveillance, immune escape data and near real-time human mobility metrics is desirable, although it may be confounded by the different

classes of neutralizing and NTD antibodies, vaccine strategies and even host heterogeneity.

That the major SARS-CoV-2 VOCs have reduced the efficacy of monoclonal antibodies and vaccines has posed serious challenges to the control of the COVID-19 pandemic. First, although vaccines can protect people infected with SARS-CoV-2 variants against severe disease, vaccine manufacturers are exploring redesigns of their products to obtain more effective protection—to eventually prevent virus transmission. Second, the suboptimal protection provided by vaccines<sup>150</sup> and the deployment of antibody-based treatments of limited or undemonstrated efficacy<sup>151</sup> has raised concerns that this would accelerate the emergence of new variants, although there is a strong argument for mass vaccination even if vaccines can only provide partial immunity<sup>152,153</sup>. Third, this has also raised the possibility that SARS-CoV-2 will become a recurrent seasonal infection<sup>154,155</sup>. Fourth, because vaccines cannot completely prevent transmission of the major variants, some NPIs such as face covering might have to be implemented to reduce transmission of the virus, as unlimited, large-scale spread of the variants would probably generate more new variants.

The genomic surveillance of SARS-CoV-2 is also facing several major challenges. First, despite this enormous endeavour, in reality only a very small proportion (around 2%) of cases have been sequenced. In addition, the majority of sequences come from a small number of countries and, remarkably, as of August 2021, around 50% of genomes have been generated in the UK and the USA, which have led the worldwide effort in this respect. By contrast, other countries with major outbreaks, such as India and Brazil, have sequenced a much smaller numbers of cases, which may cause delays in identifying variants with previously undescribed phenotypic characteristics. Therefore, it is likely that there are additional new variants that have not yet been detected given the limited genomic surveillance in a number of regions. Indeed, because the major VOCs are genetically divergent, it is possible that they have been circulating cryptically in unsampled locations, or have also emerged in individuals with a chronically infection who shed the virus for extended periods<sup>156,157</sup>. Second, the complex transmission dynamics caused by different SARS-CoV-2 variants and their continuous evolution clearly necessitate increased genomic surveillance in such a world with global connectivity and travel networks reshaped by the pandemic. Third, it is possible that recombination among viruses will also change the genetic structure of SARS-CoV-2, perhaps generating viruses with an altered phenotype. Indeed, there have already been suggestions of recombination between the Alpha and Epsilon variants in California in early 2021 (ref. <sup>158</sup>). Similarly, the potential recombination of SARS-CoV-2 and other mild human coronaviruses should not be neglected.

In summary, SARS-CoV-2 has led to an increased understanding of coronavirus evolution and the virus has entered a new evolutionary phase characterized by the frequent emergence and spread of variants that affect immune escape and reduce the efficacy of vaccines. Of particular concern is that the limited genomic surveillance in many low-income countries may cause delays in identifying variants with previously undescribed phenotypic characteristics. To contain the current and future pandemics, we urgently call for closer international cooperation, increased vaccine supply and sharing, rapid information exchange, and the establishment of both the infrastructure and trained personnel required for the effective genomic surveillance of SARS-CoV-2 and other emerging viruses.

## References

1. 1.

Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* **382**, 727–733 (2020). **One of the first papers to describe the identification of SARS-CoV-2 in Wuhan in late December 2019, providing the sequences of three full-length viral genomes and the successful isolation of the novel coronavirus.**

2. 2.

Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020). **This study describes the genomic structure and phylogenetic position of eight complete and two partial SARS-CoV-2 genome sequences obtained from samples of nine patients from different hospitals in Wuhan in late December, 2019.**

3. 3.

Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

4. 4.

Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020). **In addition to providing the early identification of SARS-CoV-2, this paper includes a description of the closest relative of SARS-CoV-2 identified to date—the bat-derived coronavirus RaTG13.**

5. 5.

Tan, W. et al. A novel coronavirus genome identified in a cluster of pneumonia cases - Wuhan, China 2019–2020. *China CDC Weekly* **2**, 61–62 (2020). **Genomic sequencing of the novel coronavirus, initially named after nCoV-19.**

6. 6.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).

7. 7.

WHO. *WHO-convened Global Study of Origins of SARS-CoV-2: China Part*. Joint WHO–China Study (WHO, 2021).

8. 8.

Hill, V. & Rambaut, A. Phylodynamic analysis of SARS-CoV-2. *Virological* <https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420> (2020).

9. 9.

Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020). **An early estimate of the epidemiological characteristics at the start of the COVID-19 outbreak, providing important evidence of human-to-human transmission from the middle of December 2019 in Wuhan, China.**

10. 10.

Holmes, E. C. et al. The origins of SARS-CoV-2: a critical review. *Cell* **184**, 4848–4856 (2021).

11. 11.

Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).

12. 12.

Wang, Q. H. et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **181**, 894–904 (2020). **Identification and structural basis of the binding of the ACE2 receptor to SARS-CoV-2.**

13. 13.

Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

14. 14.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020). **This paper describes the fundamental genomic features of SARS-CoV-2—in particular, the RBD and the furin cleavage site—and outlines the case for its zoonotic origin.**

15. 15.

Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **94**, e00127-20 (2020).

16. 16.

Garry, R. F. & Gallaher, W. R. Naturally occurring indels in multiple coronavirus spikes. *Virological* <https://virological.org/t/naturally-occurring-indels-in-multiple-coronavirus-spikes/560> (2020).

17. 17.

Li, X. et al. A furin cleavage site was discovered in the S protein of the 2019 novel coronavirus [in Chinese]. *Chin. J. Bioinform.* **18**, 103–108 (2020).

18. 18.

Wrobel, A. G. et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).

19. 19.

Johnson, B. A. et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293–299 (2021).

20. 20.

Zhou, P. et al. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **588**, E6 (2020).

21. 21.

Boni, M. F. et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* **5**, 1408–1417(2020).

22. 22.

Lam, T. T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

23. 23.

Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286–289(2020).

24. 24.

Li, X. et al. Pathogenicity, tissue tropism and potential vertical transmission of SARS-CoV-2 in Malayan pangolins. Preprint at <https://doi.org/10.1101/2020.06.22.164442> (2020).

25. 25.

Zhou, H. et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**, 2196–2203 (2020).

26. 26.

Zhou, H. et al. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* **184**, 4380–4391 (2021).

27. 27.

Murakami, S. et al. Detection and characterization of bat sarbecovirus phylogenetically related to SARS-CoV-2, Japan. *Emerg. Infect. Dis.* **26**, 3025–3029 (2020).

28. 28.

Delaune, D. et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat. Commun.* **12**, 6563 (2021).

29. 29.

Wacharapluesadee, S. et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021).

30. 30.

Temmam, S. et al. Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. Preprint at <https://doi.org/10.21203/rs.3.rs-871965/v1> (2021).

31. 31.

Bouvet, M. et al. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl Acad. Sci. USA* **109**, 9372–9377 (2012).

32. 32.

Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018).

33. 33.

Grubaugh, N. D. et al. Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **4**, 10–19 (2019).

34. 34.

Rockett, R. J. et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **26**, 1398–1404 (2020).

35. 35.

Popa, A. et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl.*

*Med.* **12**, eabe2555 (2020).

36. 36.

Oude Munnink, B. B. et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **26**, 1405–1410 (2020).

37. 37.

Miller, D. et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nat. Commun.* **11**, 5518 (2020).

38. 38.

du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021). **This milestone paper describes the genomic surveillance of SARS-CoV-2 in the UK and used an analysis of more than 50,000 genome sequences to analyse the structure of SARS-CoV-2 lineages at a fine scale.**

39. 39.

Zhong, N. S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353–1358 (2003).

40. 40.

Lu, J. et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003 (2020). **A detailed description of the genomic epidemiology of SARS-CoV-2 in Guangdong province, China, illustrating how genomic surveillance facilitated outbreak containment in China.**

41. 41.

Liu, Z. et al. Identification of common deletions in the spike protein of severe acute respiratory syndrome coronavirus 2. *J. Virol.* **94**, e00790-20 (2020).

42. 42.

Imported coronavirus variant case reported in Guangdong. *XINHUA* [http://www.xinhuanet.com/english/2021-01/03/c\\_139637931.htm](http://www.xinhuanet.com/english/2021-01/03/c_139637931.htm) (3 January 2021).

43. 43.

Guangdong reports imported coronavirus variant case. *XINHUA* [http://www.xinhuanet.com/english/2021-01/06/c\\_139646690.htm](http://www.xinhuanet.com/english/2021-01/06/c_139646690.htm) (6 January 2021).

44. 44.

Holshue, M. L. et al. First case of 2019 novel coronavirus in the United States. *New Engl. J. Med.* **382**, 929–936 (2020).

45. 45.

World Health Organization. *Coronavirus disease 2019 (COVID-19) Situation Report – 26* [https://www.who.int/docs/default-source/coronavirus/situation-reports/20200215-sitrep-26-covid-19.pdf?sfvrsn=a4cc6787\\_2](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200215-sitrep-26-covid-19.pdf?sfvrsn=a4cc6787_2) (WHO, 2020).

46. 46.

Worobey, M. et al. The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).

47. 47.

Gonzalez-Reiche, A. S. et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **369**, 297–301 (2020). **This paper analysed the early transmission dynamics of SARS-CoV-2 in New York City, highlighting viral introductions from Europe to the USA.**

48. 48.

Deng, X. et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **369**, 582–587 (2020).

49. 49.

Bedford, T. et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**, 571–575 (2020). **This paper describes the cryptic transmission of SARS-CoV-2 in the USA.**

50. 50.

Maurano, M. T. et al. Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. *Genome Res.* **30**, 1781–1788 (2020).

51. 51.

Fauver, J. R. et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* **181**, 990–996 (2020). **This study describes the long-distance domestic spread of SARS-CoV-2 in the USA.**

52. 52.

Moreno, G. K. et al. Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. *Nat. Commun.* **11**, 5558 (2020).

53. 53.

Duchene, S. et al. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* **6**, veaa061 (2020).

54. 54.

Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

55. 55.

Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* **6**, eabb5813 (2020).

56. 56.

Simmonds, P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* **5**, e00408-20 (2020).

57. 57.

Lu, G., Wang, Q. & Gao, G. F. Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* **23**, 468–478 (2015).

58. 58.

Wang, Q., Wong, G., Lu, G., Yan, J. & Gao, G. F. MERS-CoV spike protein: targets for vaccines and therapeutics. *Antiviral Res.* **133**, 165–177 (2016).

59. 59.

Korber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827 (2020). **An important paper that provided the initial evidence that the D614G substitution increased the infectivity—and therefore transmissibility—of SARS-CoV-2.**

60. 60.

Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75 (2021).

61. 61.

Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2020).

62. 62.

Zhang, L. et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* **11**, 6013 (2020).

63. 63.

Yurkovetskiy, L. et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* **183**, 739–751 (2020).

64. 64.

Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).

65. 65.

Lassaunière, R. et al. *SARS-CoV-2 Spike Mutations Arising in Danish Mink and their Spread to Humans (Working paper of SSI)*, [https://files.ssi.dk/Mink-cluster-5-short-report\\_AFO2](https://files.ssi.dk/Mink-cluster-5-short-report_AFO2) (2020).

66. 66.

Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).

67. 67.  
Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
68. 68.  
Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2020).
69. 69.  
Tegally, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARSCoV-2) lineage with multiple spike mutations in South Africa. Preprint at <https://doi.org/10.1101/2020.12.21.20248640> (2020).
70. 70.  
Faria, N. R. et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological* <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586> (2021).
71. 71.  
Naveca, F. et al. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological* <https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spikeprotein/585> (2021).
72. 72.  
O'Toole, Á. & Hill, V. SARS-CoV-2 lineages: B.1.617.2 report. [https://cov-lineages.org/global\\_report\\_B.1.617.2.html](https://cov-lineages.org/global_report_B.1.617.2.html) (2021).
73. 73.  
MacLean, O. A. et al. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol.* **19**, e3001115 (2021).
74. 74.

Martin, D. P. et al. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape. Preprint at <https://doi.org/10.1101/2021.02.23.21252268> (2021).

75. 75.

Kupferschmidt, K. Viral evolution may herald new pandemic phase. *Science* **371**, 108–109 (2021).

76. 76.

Wang, P. et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).

77. 77.

Liu, C. et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220–4236 (2021).

78. 78.

Liu, J. et al. BNT162b2-elicited neutralization of B.1.617 and other SARS-CoV-2 variants. *Nature* **596**, 273–275 (2021).

79. 79.

Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).

80. 80.

Nelson, G. et al. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. Preprint at <https://doi.org/10.1101/2021.01.13.426558> (2021).

81. 81.

Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).

82. 82.

Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383 (2021).

83. 83.

Dejnirattisai, W. et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**, 2939–2954 (2021).

84. 84.

Su, Y. C. F. et al. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. *mBio* **11**, e01610-20 (2020).

85. 85.

Young, B. E. et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**, 603–611 (2020).

86. 86.

Gong, Y. N. et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerg. Microbes Infect.* **9**, 1457–1466 (2020).

87. 87.

Du, P. et al. Genomic surveillance of COVID-19 cases in Beijing. *Nat. Commun.* **11**, 5503 (2020).

88. 88.

Pan, A. et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *J. Am. Med. Assoc.* **323**, 1915–1923 (2020).

89. 89.

Hao, X. et al. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420–424 (2020). **A comprehensive study of the transmission dynamics of COVID-19 in Wuhan, China, through time, providing important lessons learnt from the interventions in the city.**

90. 90.

Xu, X. et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat. Med.* **26**, 1193–1195 (2020).

91. 91.

Liu, A. et al. Seropositive prevalence of antibodies against SARS-CoV-2 in Wuhan, China. *JAMA Netw. Open.* **3**, e2025717 (2020).

92. 92.

Chinese Center for Disease Control and Prevention. *Scientific Understanding of the Prevalence of SARS-CoV-2: Q&A on the Results of National COVID-19 Seroepidemiological Survey in China*, [http://www.chinacdc.cn/yw\\_9324/202012/t20201228\\_223494.html](http://www.chinacdc.cn/yw_9324/202012/t20201228_223494.html) (Chinese CDC, 2020).

93. 93.

Li, Z. et al. Antibody seroprevalence in the epicenter Wuhan, Hubei, and six selected provinces after containment of the first epidemic wave of COVID-19 in China. *Lancet Reg. Health West Pac.* **8**, 100094 (2021).

94. 94.

Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).

95. 95.

Niehus, R., De Salazar, P. M., Taylor, A. R. & Lipsitch, M. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *Lancet Infect. Dis.* **20**, 803–808 (2020).

96. 96.

Cao, S. Y. et al. Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China. *Nat. Commun.* **11**, 5917 (2020).

97. 97.

Jia, J. S. et al. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).

98. 98.

Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).

99. 99.

Lai, S. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413 (2020). **This study quantified the effects of various non-pharmaceutical interventions and their timings on COVID-19, providing early evidence that informed response efforts around the world.**

100. 100.

Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).

101. 101.

Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, 793–802 (2020).

102. 102.

Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

103. 103.

Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* **395**, 1382–1393 (2020).

104. 104.

Zhang, Z.-B. et al. Countries of origin of imported COVID-19 cases into China and measures to prevent onward transmission. *J. Travel Med.* **27**, taaa139 (2020).

105. 105.

Bai, L. et al. Controlling COVID-19 transmission due to contaminated imported frozen food and food packaging. *China CDC Wkly* **3**, 30–33 (2021).

106. 106.

Pang, X. et al. Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing. *Natl Sci. Rev.* **7**, 1861–1864 (2020).

107. 107.

Xing, Y., Wong, G. W. K., Ni, W., Hu, X. & Xing, Q. Rapid response to an outbreak in Qingdao, China. *N. Engl. J. Med.* **383**, e129 (2020).

108. 108.

Bogoch, I. I. et al. Potential for global spread of a novel coronavirus from China. *J. Travel Med.* **27**, taaa0111 (2020).

109. 109.

Lai, S., Bogoch, I. I., Watts, A., Khan, K. & Tatem, A. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China. *WorldPop* <https://www.worldpop.org/events/china> (2020).

110. 110.

Yang, J. et al. Uncovering two phases of early intercontinental COVID-19 transmission dynamics. *J. Travel Med.* **27**, taaa200 (2020).

111. 111.

Pullano, G. et al. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Euro Surveill.* **25**, 2000057 (2020).

112. 112.

Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020).

113. 113.

Wells, C. R. et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl Acad. Sci. USA* **117**, 7504–7509 (2020).

114. 114.

Russell, T. W. et al. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *Lancet Public Health* **6**, e12–e20 (2021).

115. 115.

Devi, S. COVID-19 resurgence in Iran. *Lancet* **395**, 1896 (2020).

116. 116.

Giordano, G. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855–860 (2020).

117. 117.

Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).

118. 118.

Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).

119. 119.

Bushman, D. et al. Detection and genetic characterization of community-based SARS-CoV-2 infections — New York City, March 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 918–922 (2020).

120. 120.

Roszen, L. M., Branum, A. M., Ahmad, F. B., Sutton, P. & Anderson, R. N. Excess deaths associated with COVID-19, by age and race and ethnicity — United States, January 26–October 3, 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 1522–1527 (2020).

121. 121.

Candido, D. S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**, 1255–1260 (2020).

122. 122.

Silveira, M. F. et al. Population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. *Nat. Med.* **26**, 1196–1199 (2020).

123. 123.

Acharya, R. & Porwal, A. A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. *Lancet Glob. Health* **8**, e1142–e1151 (2020).

124. 124.

Laxminarayan, R. et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020).

125. 125.

Massinga Loembé, M. et al. COVID-19 in Africa: the spread and response. *Nat. Med.* **26**, 999–1003 (2020).

126. 126.

Gilbert, M. et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet* **395**, 871–877 (2020).

127. 127.

Pullano, G. et al. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* **590**, 134–139 (2021).

128. 128.

Rice, B. L. et al. Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa. *Nat. Med.* **27**, 447–453 (2021).

129. 129.

Salyer, S. J. et al. The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. *Lancet* **397**, 1265–1275 (2021).

130. 130.

Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).

131. 131.

Brauner, J. M. et al. Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, eabd9338 (2020). **A chronological and global dataset was constructed to compare the effectiveness of different NPIs in reducing COVID-19 transmission among countries during the first wave of the COVID-19 pandemic.**

132. 132.

Haug, N. et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).

133. 133.

Dehning, J. et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, eabb9789 (2020).

134. 134.

Baker, R. E., Yang, W. C., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020).

135. 135.

Han, E. et al. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *Lancet* **396**, 1525–1534 (2020).

136. 136.

Badr, H. S. et al. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 1247–1254 (2020).

137. 137.

Perkins, T. A. et al. Estimating unobserved SARS-CoV-2 infections in the United States. *Proc. Natl Acad. Sci. USA* **117**, 22597–22602 (2020).

138. 138.

Centers for Disease Control and Prevention. *COVID Data Tracker*.  
[https://covid.cdc.gov/covid-data-tracker/#cases\\_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days) (US CDC, 2020).

139. 139.

The Ministry of Health, Brazil. *Coronavirus, Brazil*. <https://covid.saude.gov.br/> (2021).

140. 140.

Lemey, P. et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. **Nature** **595**, 713–717 (2021).

141. 141.

Hodcroft, E. B. et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. **Nature** **595**, 707–712 (2021).

142. 142.

European Centre for Disease Prevention and Control. *COVID-19*.  
<https://www.ecdc.europa.eu/en/covid-19-pandemic> (ECDC, 2021).

143. 143.

Kupferschmidt, K. & Wadman, M. Delta variant triggers new phase in the pandemic. **Science** **372**, 1375–1376 (2021).

144. 144.

World Health Organization. *Weekly epidemiological update on COVID-19 – 29 June 2021*. <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-june-2021> (WHO, 2021).

145. 145.

WHO. *Tracking SARS-CoV-2 variants*,  
<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (2021).

146. 146.

Centers for Disease Control and Prevention. *SARS-CoV-2 Variant Classifications and Definitions*. <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html> (US CDC, 2021).

147. 147.

Harvey, W. T. et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).

148. 148.

Li, Q. et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294 (2020).

149. 149.

Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 (2021).

150. 150.

Saad-Roy, C. M. et al. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science* **372**, 363–370 (2021).

151. 151.

Kemp, S. A. et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277–282 (2021).

152. 152.

Cobey, S., Larremore, D. B., Grad, Y. H. & Lipsitch, M. Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nat. Rev. Immunol.* **21**, 330–335 (2021).

153. 153.

Hanage, W. P. & Russell, C. A. Partial immunity and SARS-CoV-2 mutations. *Science* **372**, 354 (2021).

154. 154.

Murray, C. J. L. & Piot, P. The potential future of the COVID-19 pandemic: will SARS-CoV-2 become a recurrent seasonal infection? *J. Am. Med. Assoc.* **325**, 1249–1250 (2021).

155. 155.

Phillips, N. The coronavirus is here to stay—here's what that means. *Nature* **590**, 382–384 (2021).

156. 156.

Avanzato, V. A. et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell* **183**, 1901–1912 (2020).

157. 157.

Aydillo, T. et al. Shedding of viable SARS-CoV-2 after immunosuppressive therapy for cancer. *N. Engl. J. Med.* **383**, 2586–2588 (2020).

158. 158.

Lawton, G. Exclusive: Two variants have merged into heavily mutated coronavirus. *New Scientist* <https://www.newscientist.com/article/2268014-exclusive-two-variants-have-merged-into-heavily-mutated-coronavirus/> (16 February 2021).

159. 159.

Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

160. 160.

GISAID. UK reports New Variant, termed VUI 202012/01. <https://www.gisaid.org/references/gisaid-in-the-news/uk-reports-new-variant-termied-vui-20201201/> (2021).

161. 161.

Song, S. et al. The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoVR. *Genom. Proteom. Bioinform.* **18**, 749–759 (2020).

162. 162.

Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

163. 163.

Davies, N. G. et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).

164. 164.

Frampton, D. et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect. Dis.* **21**, 1246–1256 (2021).

165. 165.

Graham, M. S. et al. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study. *Lancet Public Health* **6**, e335–e345 (2021).

166. 166.

World Health Organization. *Weekly epidemiological update on COVID-19 – 10 August 2021*. <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2021> (WHO, 2021).

167. 167.

European Centre for Disease Prevention and Control. *Emergence of SARS-CoV-2 B.1.617 variants in India and situation in the EU/EEA*. <https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-sars-cov-2-b1617-variants> (ECDC, 2021).

168. 168.

Delta Plus: Key things to know about new coronavirus variant. *The Economic Times* <https://economictimes.indiatimes.com/news/et-explains/delta-plus-key-things-to-know-about-new-coronavirus-variant-in-india/articleshow/83739996.cms> (24 June 2021).

169. 169.

Gu, H. et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603–1607 (2020).

## Acknowledgements

We acknowledge the efforts of the World Health Organization in sharing the COVID-19 Weekly Epidemiological Update, the researchers that are part of the cov-lineages.org team (<https://cov-lineages.org/>) in assembling the records for new strains, the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) for collating the COVID-19 case data ([github.com/CSSEGISandData/COVID-19](https://covid19.jhu.edu/data/)), and Google for producing and sharing the COVID-19 Community Mobility Reports ([www.google.com/covid19/mobility/](https://www.google.com/covid19/mobility/)). We thank the researchers who generated and shared the sequencing data from the GISAID and NCBI GenBank databases; T. Hu, P. Wang, X. Yao and H. Song for helping to produce the figures; and E. C. Holmes, A. Tatem and J. Floyd for commenting on the manuscript. This work was supported by Key Research and Development Project of Shandong province (grant nos. 2020SFXY01 and 2020SFXY08), the National Key Research and Development Programme of China (grant no. 2020YFC0840800), the National Science and Technology Major Project (Grant no. 2018ZX10101004-002 and 2016ZX10004222-009), the National Natural Science Fund of China (81773498), the Academic Promotion Programme of Shandong First Medical University (2019QL006), and the Bill & Melinda Gates Foundation (INV-024911). W.S. was supported by the Taishan Scholars Program of Shandong Province (ts201511056).

## Author information

### Author notes

1. These authors contributed equally: Juan Li, Shengjie Lai

### Affiliations

1. School of Public Health, Shandong First Medical University & Shandong Academy of Medical Sciences, Tai'an, China

Juan Li & Weifeng Shi

2. Key Laboratory of Etiology and Epidemiology of Emerging Infectious Diseases in the Universities of Shandong, Shandong First Medical University & Shandong Academy of Medical Sciences, Tai'an, China

Juan Li & Weifeng Shi

3. WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK

Shengjie Lai

4. National Institute for Viral Disease Control and Prevention, China CDC, Beijing, China

George F. Gao

5. CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology,, Chinese Academy of Sciences, Beijing, China

George F. Gao

6. Center for Influenza Research and Early-warning (CASCIRE), CAS-TWAS Center of Excellence for Emerging Infectious Diseases (CEEID), Chinese Academy of Sciences, Beijing, China

George F. Gao

## Contributions

W.S. conceived the study. J.L. performed phylogenetic analysis and homology modelling. S.L. conducted the literature review on the global spread of SARS-CoV-2 and VOCs, and collected, analysed and visualized the data of the case numbers, VOC reports and human mobility, using publicly available data resources. W.S., J.L. and S.L. wrote the first draft of the manuscript. W.S. and G.F.G. proofread the manuscript and the pre-submission inquiry.

## Corresponding author

Correspondence to [Weifeng Shi](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Malik Peiris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 The date of the earliest sequence of VOC detected in each country, territory or area.](#)

**a**, VOC Alpha (lineage B.1.1.7). **b**, VOC Beta (lineage B.1.351). **c**, VOC Gamma (lineage P.1). **d**, VOC Delta (B.1.617.2). Reporting date of the earliest sequence of each VOC used records published at <https://cov-lineages.org/>, as of 30 July 2021, derived from publicly available sequence data in GISAID, shared by international sequencing efforts. The areas without data are shown in grey.

## Supplementary information

### [Supplementary Table 1](#)

The top 50 ranked countries with the highest number of cases of COVID-19 reported as of 8 August 2021 and the available Google mobility data.

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Li, J., Lai, S., Gao, G.F. *et al.* The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **600**, 408–418 (2021). <https://doi.org/10.1038/s41586-021-04188-6>

- Received: 28 January 2021
- Accepted: 26 October 2021
- Published: 08 December 2021
- Issue Date: 16 December 2021

- DOI: <https://doi.org/10.1038/s41586-021-04188-6>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04188-6>

| [Section menu](#) | [Main menu](#) |

- Perspective
- [Published: 15 December 2021](#)

# Biogeochemical extremes and compound events in the ocean

- [Nicolas Gruber](#) ORCID: [orcid.org/0000-0002-2085-2310](https://orcid.org/0000-0002-2085-2310)<sup>1</sup>,
- [Philip W. Boyd](#) ORCID: [orcid.org/0000-0001-7850-1911](https://orcid.org/0000-0001-7850-1911)<sup>2</sup>,
- [Thomas L. Frölicher](#) ORCID: [orcid.org/0000-0003-2348-7854](https://orcid.org/0000-0003-2348-7854)<sup>3,4</sup> &
- [Meike Vogt](#)<sup>1</sup>

[Nature](#) volume 600, pages 395–407 (2021)

- 1174 Accesses
- 97 Altmetric
- [Metrics details](#)

## Subjects

- [Climate-change ecology](#)
- [Ocean sciences](#)

## Abstract

The ocean is warming, losing oxygen and being acidified, primarily as a result of anthropogenic carbon emissions. With ocean warming, acidification and deoxygenation projected to increase for decades, extreme events, such as marine heatwaves, will intensify, occur more often, persist for longer periods of time and extend over larger regions. Nevertheless, our understanding of oceanic extreme events that are associated with warming,

low oxygen concentrations or high acidity, as well as their impacts on marine ecosystems, remains limited. Compound events—that is, multiple extreme events that occur simultaneously or in close sequence—are of particular concern, as their individual effects may interact synergistically. Here we assess patterns and trends in open ocean extremes based on the existing literature as well as global and regional model simulations. Furthermore, we discuss the potential impacts of individual and compound extremes on marine organisms and ecosystems. We propose a pathway to improve the understanding of extreme events and the capacity of marine life to respond to them. The conditions exhibited by present extreme events may be a harbinger of what may become normal in the future. As a consequence, pursuing this research effort may also help us to better understand the responses of marine organisms and ecosystems to future climate change.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

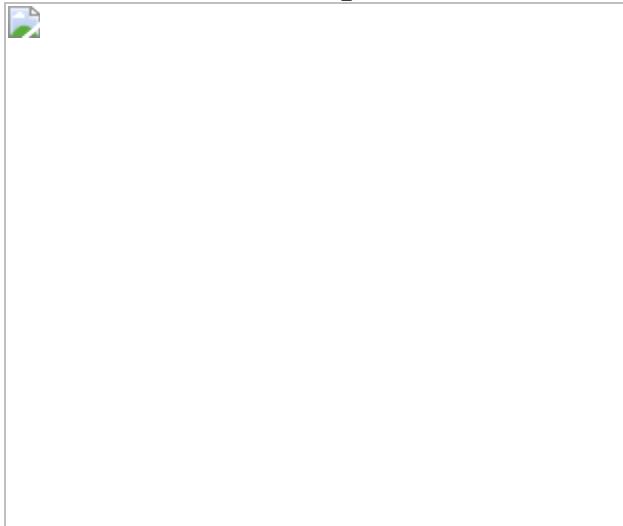
[Rent or Buy](#)

All prices are NET prices.

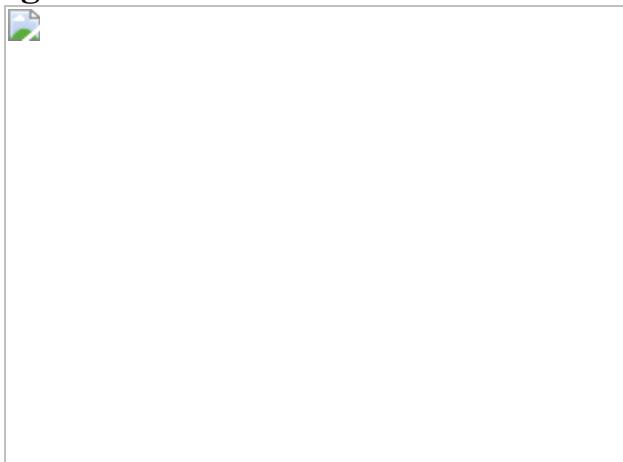
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

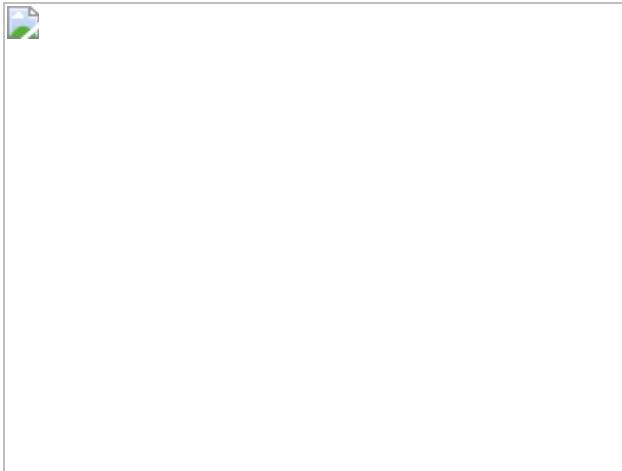
**Fig. 1: Drivers of marine heatwaves and ocean biogeochemical extremes and their impact.**



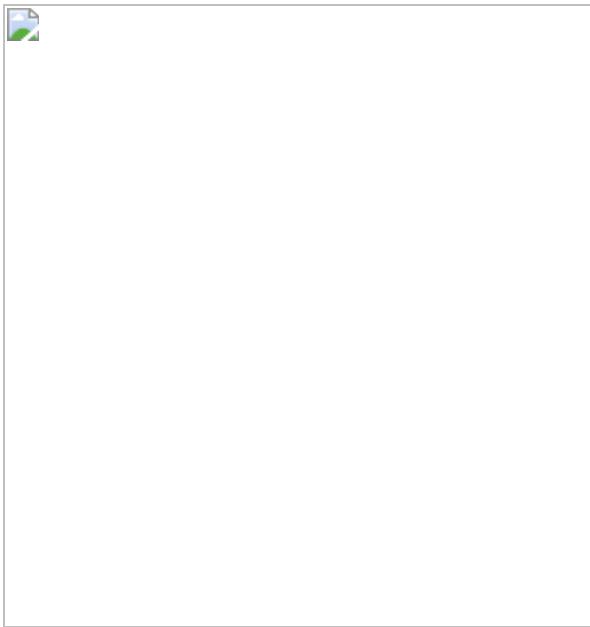
**Fig. 2: Global pattern and trends of marine heatwaves and biogeochemical extremes.**



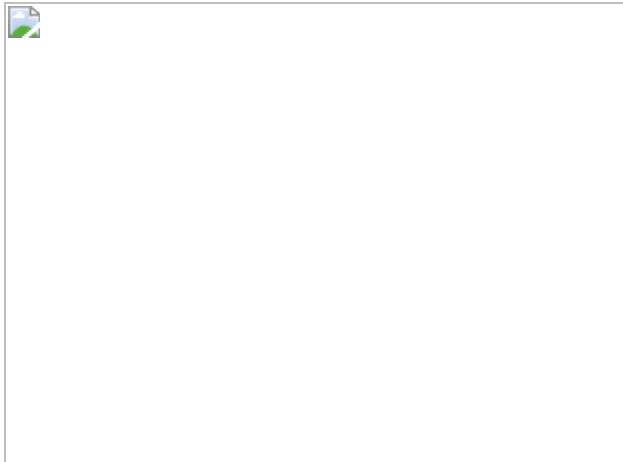
**Fig. 3: Idealized depictions of the mechanisms giving rise to extreme conditions in temperature, acidity and oxygen.**



**Fig. 4: Illustration of the temporal evolution and four-dimensional nature of ocean extremes using model simulation results for the 2013–2015 Blob.**



**Fig. 5: Conceptual framework to understand the impact of marine biogeochemical extremes and heatwaves on the functioning of marine life—from physiology through to ecosystems.**



## Data availability

The model results presented in the paper are available through the ETH Research Collection (<https://www.research-collection.ethz.ch>, <https://doi.org/10.3929/ethz-b-000501082>).

## References

1. 1.

Gruber, N. Warming up, turning sour, losing breath: ocean biogeochemistry under global change. *Philos. Trans. A Math. Phys. Eng. Sci.* **369**, 1980–1996 (2011). **Identifies the potential synergistic threat to marine ecosystems resulting from ocean warming, deoxygenation and acidification.**

2. 2.

Gattuso, J.-P. et al. Contrasting futures for ocean and society from different anthropogenic CO<sub>2</sub> emissions scenarios. *Science* **349**, aac4722 (2015).

3. 3.

Breitburg, D. et al. Declining oxygen in the global ocean and coastal waters. *Science* **359**, eaam7240 (2018).

4. 4.

Cheng, L. et al. Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.* **3**, e1601545 (2017).

5. 5.

Friedlingstein, P. et al. Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).

6. 6.

Keeling, R. F., Kortzinger, A. & Gruber, N. Ocean deoxygenation in a warming world. *Ann. Rev. Mar. Sci.* **2**, 199–229 (2010).

7. 7.

Li, G. et al. Increasing ocean stratification over the past half-century. *Nat. Clim. Change* **10**, 1116–1123 (2020).

8. 8.

Sallée, J. B. et al. Summertime increases in upper-ocean stratification and mixed-layer depth. *Nature* **591**, 592–598 (2021).

9. 9.

Sarmiento, J. L. & Gruber, N. *Ocean Biogeochemical Dynamics* (Princeton Univ. Press, 2006).

10. 10.

Mikaloff Fletcher, S. E. et al. Inverse estimates of anthropogenic CO<sub>2</sub> uptake, transport, and storage by the ocean. *Glob. Biogeochem. Cycles* **20**, GB2002 (2006).

11. 11.

Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. Ocean acidification: the other CO<sub>2</sub> problem. *Ann. Rev. Mar. Sci.* **1**, 169–192 (2009).

12. 12.

Doney, S. C., Busch, D. S., Cooley, S. R. & Kroeker, K. J. The impacts of ocean acidification on marine ecosystems and reliant human communities. *Annu. Rev. Environ. Resour.* **45**, 83–112 (2020).

13. 13.

Kroeker, K. J., Kordas, R. L., Crim, R. N. & Singh, G. G. Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. *Ecol. Lett.* **13**, 1419–1434 (2010).

14. 14.

Mollica, N. R. et al. Ocean acidification affects coral growth by reducing skeletal density. *Proc. Natl Acad. Sci. USA* **115**, 1754–1759 (2018).

15. 15.

Bopp, L. et al. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences* **10**, 6225–6245 (2013).

16. 16.

Kwiatkowski, L. et al. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences* **17**, 3439–3470 (2020).

17. 17.

Seneviratne, S. I. et al. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) 109–230 (Cambridge Univ. Press, 2012).

18. 18.

Lavell, A. et al. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) 25–64 (2012).

19. 19.

Parmesan, C., Root, T. L. & Willig, M. R. Impacts of extreme weather and climate on terrestrial biota. *Bull. Am. Meteor. Soc.* **81**, 443–450 (2000).

20. 20.

Smith, M. D. An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research. *J. Ecol.* **99**, 656–663 (2011).

21. 21.

Oliver, E. C. J. et al. Marine heatwaves. *Ann. Rev. Mar. Sci.* **13**, 313–342 (2021).

22. 22.

Frölicher, T. L., Fischer, E. M. & Gruber, N. Marine heatwaves under global warming. *Nature* **560**, 360–364 (2018). **Quantifies the future evolution of marine heatwaves under different climate scenarios and their attribution to climate change.**

23. 23.

Oliver, E. C. J. et al. Longer and more frequent marine heatwaves over the past century. *Nat. Commun.* **9**, 1324 (2018). **Highlights the strong**

**increase in the occurrence and intensity of marine heatwaves in recent decades.**

24. 24.

Oliver, E. C. J. et al. Projected marine heatwaves in the 21st century and the potential for ecological impact. *Front. Mar. Sci.* **6**, 734 (2019).

25. 25.

Benedetti-Cecchi, L. Complex networks of marine heatwaves reveal abrupt transitions in the global ocean. *Sci. Rep.* **11**, 1739 (2021).

26. 26.

Smale, D. A. et al. Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat. Clim. Change* **9**, 306–312 (2019). **Assesses the potential for global ocean ecosystem impacts of marine heatwaves.**

27. 27.

Wernberg, T. et al. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat. Clim. Change* **3**, 78–82 (2012). **Demonstrates marked ocean ecosystem changes in response to a heatwave.**

28. 28.

Hobday, A. J. et al. A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.* **141**, 227–238 (2016).

29. 29.

Holbrook, N. J. et al. A global assessment of marine heatwaves and their drivers. *Nat. Commun.* **10**, 2624 (2019).

30. 30.

Huang, B. et al. Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: upgrades and intercomparisons. *J. Clim.* **28**, 911–930 (2015).

31. 31.

Gentemann, C. L., Fewings, M. R. & García-Reyes, M. Satellite sea surface temperatures along the west coast of the United States during the 2014–2016 northeast Pacific marine heat wave. *Geophys. Res. Lett.* **44**, 312–319 (2017).

32. 32.

Cavole, L. et al. Biological impacts of the 2013–2015 warm-water anomaly in the northeast Pacific: winners, losers, and the future. *Oceanography* **29**, 273–285 (2016). **A synthesis of the ecosystem impacts of the 2013–2015 Blob heatwave.**

33. 33.

Di Lorenzo, E. & Mantua, N. Multi-year persistence of the 2014/15 north Pacific marine heatwave. *Nat. Clim. Change* **6**, 1042–1047 (2016).

34. 34.

Brodeur, R. D., Auth, T. D. & Phillips, A. J. Major shifts in pelagic micronekton and macrozooplankton community structure in an upwelling ecosystem related to an unprecedented marine heatwave. *Front. Mar. Sci.* **6**, 212 (2019).

35. 35.

Cheung, W. W. L. & Frölicher, T. L. Marine heatwaves exacerbate climate change impacts for fisheries in the northeast Pacific. *Sci. Rep.* **10**, 6678 (2020).

36. 36.

Laufkötter, C., Zscheischler, J. & Frölicher, T. L. High-impact marine heatwaves attributable to human-induced global warming. *Science* **369**, 1621–1625 (2020).

37. 37.

Hauri, C., Gruber, N., McDonnell, A. M. P. & Vogt, M. The intensity, duration, and severity of low aragonite saturation state events on the California continental shelf. *Geophys. Res. Lett.* **40**, 3424–3428 (2013). **Models the evolution of ocean-acidification-related extremes in the California Current System.**

38. 38.

Burger, F. A., John, J. G. & Frölicher, T. L. Increase in ocean acidity variability and extremes under increasing atmospheric CO<sub>2</sub>. *Biogeosciences* **17**, 4633–4662 (2020).

39. 39.

Leonard, M. et al. A compound event framework for understanding extreme impacts. *Wiley Interdiscip. Rev. Clim. Change* **5**, 113–128 (2014).

40. 40.

Zscheischler, J. et al. Future climate risk from compound events. *Nat. Clim. Change* **8**, 469–477 (2018).

41. 41.

Le Grix, N., Zscheischler, J., Laufkötter, C., Rousseaux, C. S. & Frölicher, T. L. Compound high-temperature and low-chlorophyll extremes in the ocean over the satellite period. *Biogeosciences* **18**, 2119–2137 (2021).

42. 42.

Boyd, P. W. & Brown, C. J. Modes of interactions between environmental drivers and marine biota. *Front. Mar. Sci.* **2**, 9 (2015).

43. 43.

Trisos, C. H., Merow, C. & Pigot, A. L. The projected timing of abrupt ecological disruption from climate change. *Nature* **580**, 496–501 (2020).

44. 44.

Limburg, K. E., Breitburg, D., Swaney, D. P. & Jacinto, G. Ocean deoxygenation: a primer. *One Earth* **2**, 24–29 (2020).

45. 45.

Dunne, J. P. et al. GFDL's ESM2 global coupled climate-carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics. *J. Clim.* **26**, 2247–2267 (2013).

46. 46.

Dunne, J. P. et al. GFDL's ESM2 global coupled climate–carbon earth system models. Part I: physical formulation and baseline simulation characteristics. *J. Clim.* **25**, 6646–6665 (2012).

47. 47.

Allen, M. R. et al. *An IPCC Special Report on the Impacts of Global Warming of 1.5 °C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change* (eds. Masson-Delmotte, V. et al.) 49–91 (IPCC, 2018).

48. 48.

Frölicher, T. L. & Laufkötter, C. Emerging risks from marine heat waves. *Nat. Commun.* **9**, 650 (2018).

49. 49.

Pilo, G. S., Holbrook, N. J., Kiss, A. E. & Hogg, A. M. C. sensitivity of marine heatwave metrics to ocean model resolution. *Geophys. Res. Lett.* **46**, 14604–14612 (2019).

50. 50.

Schlegel, R. W., Oliver, E. C. J., Hobday, A. J. & Smit, A. J. Detecting marine heatwaves with sub-optimal data. *Front. Mar. Sci.* **6**, 737 (2019).

51. 51.

Hobday, A. et al. Categorizing and naming marine heatwaves. *Oceanography* **31**, 162–173 (2018).

52. 52.

Alexander, M. A. et al. The atmospheric bridge: the influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Clim.* **15**, 2205–2231 (2002).

53. 53.

Amaya, D. J., Miller, A. J., Xie, S. P. & Kosaka, Y. Physical drivers of the summer 2019 north Pacific marine heatwave. *Nat. Commun.* **11**, 1903 (2020).

54. 54.

Negrete-García, G., Lovenduski, N. S., Hauri, C., Krumhardt, K. M. & Lauvset, S. K. Sudden emergence of a shallow aragonite saturation horizon in the Southern Ocean. *Nat. Clim. Change* **9**, 313–317 (2019).

55. 55.

Schaeffer, A. & Roughan, M. Subsurface intensification of marine heatwaves off southeastern Australia: the role of stratification and

local winds. *Geophys. Res. Lett.* **44**, 5025–5033 (2017).

56. 56.

Jackson, J. M., Johnson, G. C., Dosser, H. V. & Ross, T. Warming from recent marine heatwave lingers in deep British Columbia fjord. *Geophys. Res. Lett.* **45**, 9757–9764 (2018).

57. 57.

Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M. & Riser, S. C. Subsurface evolution and persistence of marine heatwaves in the northeast Pacific. *Geophys. Res. Lett.* **47**, e2020GL090548 (2020).

58. 58.

Deutsch, C., Brix, H., Ito, T., Frenzel, H. & Thompson, L. Climate-forced variability of ocean hypoxia. *Science* **333**, 336–339 (2011).

59. 59.

Frenger, I. et al. Biogeochemical role of subsurface coherent eddies in the ocean: tracer cannonballs, hypoxic storms, and microbial stewpots? *Glob. Biogeochem. Cycles* **32**, 226–249 (2018).

60. 60.

Schütte, F. et al. Characterization of ‘dead-zone’ eddies in the eastern tropical north Atlantic. *Biogeosciences* **13**, 5865–5881 (2016).

61. 61.

Lauvset, S. K. et al. Processes driving global interior ocean pH distribution. *Glob. Biogeochem. Cycles* **34**, e2019GB006229 (2020).

62. 62.

Gaube, P., Chelton, D. B., Strutton, P. G. & Behrenfeld, M. J. Satellite observations of chlorophyll, phytoplankton biomass, and Ekman

pumping in nonlinear mesoscale eddies. *J. Geophys. Res. Oceans* **118**, 6349–6370 (2013).

63. 63.

Frenger, I., Münnich, M., Gruber, N. & Knutti, R. Southern Ocean eddy phenomenology. *J. Geophys. Res. Oceans* **120**, 7413–7449 (2015).

64. 64.

Hauss, H. et al. Dead zone or oasis in the open ocean? Zooplankton distribution and migration in low-oxygen modewater eddies. *Biogeosciences* **13**, 1977–1989 (2016).

65. 65.

Gruber, N. et al. Rapid progression of ocean acidification in the California Current System. *Science* **337**, 220–223 (2012).

66. 66.

Santora, J. A. et al. Habitat compression and ecosystem shifts as potential links between marine heatwave and record whale entanglements. *Nat. Commun.* **11**, 536 (2020).

67. 67.

Bond, N. A., Cronin, M. F., Freeland, H. & Mantua, N. Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophys. Res. Lett.* **42**, 3414–3420 (2015).

68. 68.

Peterson, W. T., Bond, N. A. & Robert, M. The Blob (part three): going, going, gone? *PICES Press* **24**, 46–48 (2016).

69. 69.

Frischknecht, M., Münnich, M. & Gruber, N. Local atmospheric forcing driving an unexpected California Current System response during the 2015–2016 El Niño. *Geophys. Res. Lett.* **44**, 304–311 (2017).

70. 70.

Pörtner, H. O. & Knust, R. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* **315**, 95–97 (2007).

71. 71.

Penn, J. L., Deutsch, C., Payne, J. L. & Sperling, E. A. Temperature-dependent hypoxia explains biogeography and severity of end-Permian marine mass extinction. *Science* **362**, eaat1327 (2018).

72. 72.

Dahlke, F. T., Wohlrab, S., Butzin, M. & Pörtner, H.-O. Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science* **369**, 65–70 (2020).

73. 73.

Pörtner, H. O. Ecosystem effects of ocean acidification in times of ocean warming: a physiologist's view. *Mar. Ecol. Progr. Ser.* **373**, 203–217 (2008).

74. 74.

Stillman, J. H. Heat waves, the new normal: summertime temperature extremes will impact animals, ecosystems, and human communities. *Physiology* **34**, 86–100 (2019).

75. 75.

Straub, S. C. et al. Resistance, extinction, and everything in between—the diverse responses of seaweeds to marine heatwaves. *Front. Mar.*

*Sci.* **6**, 763 (2019).

76. 76.

Hughes, T. P. et al. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* **359**, 80–83 (2018). **Demonstrates the global-scale impact of marine heatwaves on warm-water corals.**

77. 77.

Donovan, M. K. et al. Local conditions magnify coral loss following marine heatwaves. *Science* **372**, 977–980 (2021).

78. 78.

Klein, S. G. et al. Projecting coral responses to intensifying marine heatwaves under ocean acidification. *Glob. Change Biol.* <https://doi.org/10.1111/gcb.15818> (2021).

79. 79.

McMahon, B. R. Physiological responses to oxygen depletion in intertidal animals. *Am. Zool.* **28**, 39–53 (1988).

80. 80.

Kroeker, K. J. et al. Ecological change in dynamic environments: accounting for temporal environmental variability in studies of ocean change biology. *Glob. Change Biol.* **26**, 54–67 (2020).

81. 81.

Hofmann, G. E. et al. High-frequency dynamics of ocean pH: a multi-ecosystem comparison. *PLoS ONE* **6**, e28983 (2011).

82. 82.

Spisla, C. et al. Extreme levels of ocean acidification restructure the plankton community and biogeochemistry of a temperate coastal ecosystem: a mesocosm study. *Front. Mar. Sci.* **7**, 611157 (2021).

83. 83.

Engström-Öst, J. et al. Eco-physiological responses of copepods and pteropods to ocean warming and acidification. *Sci. Rep.* **9**, 4748 (2019).

84. 84.

Bednaršek, N. et al. El Niño-related thermal stress coupled with upwelling-related ocean acidification negatively impacts cellular to population-level responses in pteropods along the California current system with implications for increased bioenergetic costs. *Front. Mar. Sci.* **5**, 486 (2018). **Shows the impact of a compound event on pteropods, a keystone zooplankton species in many marine ecosystems.**

85. 85.

Calderón-Liévanos, S. et al. Survival and respiration of green abalone (*Haliotis fulgens*) facing very short-term marine environmental extremes. *Mar. Freshw. Behav. Physiol.* **52**, 1–15 (2019).

86. 86.

Mieszkowska, N., Burrows, M. T., Hawkins, S. J. & Sugden, H. Impacts of pervasive climate change and extreme events on rocky intertidal communities: evidence from long-term data. *Front. Mar. Sci.* **8**, 642764 (2021).

87. 87.

Nielsen, J. M. et al. Responses of ichthyoplankton assemblages to the recent marine heatwave and previous climate fluctuations in several

northeast Pacific marine ecosystems. *Glob. Chang. Biol.* **27**, 506–520 (2020).

88. 88.

Wernberg, T. et al. Climate-driven regime shift of a temperate marine ecosystem. *Science* **353**, 169–172 (2016).

89. 89.

Garrabou, J. et al. Mass mortality in northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. *Glob. Change Biol.* **15**, 1090–1103 (2009).

90. 90.

Darling, E. S., McClanahan, T. R. & Côté, I. M. Life histories predict coral community disassembly under multiple stressors. *Glob. Change Biol.* **19**, 1930–1940 (2013).

91. 91.

Ateweberhan, M., McClanahan, T. R., Graham, N. A. J. & Sheppard, C. R. C. Episodic heterogeneous decline and recovery of coral cover in the Indian Ocean. *Coral Reefs* **30**, 739–752 (2011).

92. 92.

Weitzman, B. et al. Changes in rocky intertidal community structure during a marine heatwave in the Northern Gulf of Alaska. *Front. Mar. Sci.* **8**, 556820 (2021).

93. 93.

Samuels, T., Ryneanson, T. A. & Collins, S. Surviving heatwaves: thermal experience predicts life and death in a Southern Ocean Diatom. *Front. Mar. Sci.* **8**, 600343 (2021).

94. 94.

Vajedsamiei, J., Wahl, M., Schmidt, A. L., Yazdanpanahan, M. & Pansch, C. The higher the needs, the lower the tolerance: extreme events may select ectotherm recruits with lower metabolic demand and heat sensitivity. *Front. Mar. Sci.* **8**, 660427 (2021).

95. 95.

Bindoff, N. L. et al. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (eds. Pörtner, H.-O. et al.) Ch. 5 (IPCC, 2021).

96. 96.

IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C. B. et al.) Ch. 6 (Cambridge Univ. Press, 2014).

97. 97.

Harvey, B. P., Gwynn-Jones, D. & Moore, P. J. Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. *Ecol. Evol.* **3**, 1016–1030 (2013).

98. 98.

Gunderson, A. R., Armstrong, E. J. & Stillman, J. H. Multiple stressors in a changing world: the need for an improved perspective on physiological responses to the dynamic marine environment. *Ann. Rev. Mar. Sci.* **8**, 357–378 (2016).

99. 99.

Seifert, M., Rost, B., Trimborn, S. & Hauck, J. Meta-analysis of multiple driver effects on marine phytoplankton highlights modulating role of pCO<sub>2</sub>. *Glob. Change Biol.* **26**, 6787–6804 (2020).

100. 100.

Sampaio, E. et al. Impacts of hypoxic events surpass those of future ocean warming and acidification. *Nat. Ecol. Evol.* **5**, 311–321 (2021).

101. 101.

Bernhardt, J. R., O'Connor, M. I., Sunday, J. M. & Gonzalez, A. Life in fluctuating environments: adaptation to changing environments. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190454 (2020).

102. 102.

Somero, G. N. The cellular stress response and temperature: function, regulation, and evolution. *J. Exp. Zool. A Ecol. Integr. Physiol.* **333**, 379–397 (2020).

103. 103.

Fordyce, A. J., Ainsworth, T. D., Heron, S. F. & Leggat, W. Marine heatwave hotspots in coral reef environments: physical drivers, ecophysiological outcomes and impact upon structural complexity. *Front. Mar. Sci.* **6**, 498 (2019).

104. 104.

Krueger, T. et al. Antioxidant plasticity and thermal sensitivity in four types of *Symbiodinium* sp. *J. Phycol.* **50**, 1035–1047 (2014).

105. 105.

Reusch, T. B. H. & Boyd, P. W. Experimental evolution meets marine phytoplankton. *Evolution* **67**, 1849–1859 (2013).

106. 106.

Schlüter, L., Lohbeck, K. T., Gröger, J. P., Riebesell, U. & Reusch, T. B. H. Long-term dynamics of adaptive evolution in a globally important phytoplankton species to ocean acidification. *Sci. Adv.* **2**, e1501660 (2016).

107. 107.

Schlüter, L. et al. Adaptation of a globally important coccolithophore to ocean warming and acidification. *Nat. Clim. Change* **4**, 1024–1030 (2014).

108. 108.

Aranguren-Gassis, M., Kremer, C. T., Klausmeier, C. A. & Litchman, E. Nitrogen limitation inhibits marine diatom adaptation to high temperatures. *Ecol. Lett.* **22**, 1860–1869 (2019).

109. 109.

Dam, H. G. Evolutionary adaptation of marine zooplankton to global change. *Ann. Rev. Mar. Sci.* **5**, 349–370 (2013).

110. 110.

Hinder, S. L. et al. Multi-decadal range changes vs. thermal adaptation for north east Atlantic oceanic copepods in the face of climate change. *Glob. Change Biol.* **20**, 140–146 (2014).

111. 111.

Antell, G. S., Fenton, I. S., Valdes, P. J. & Saupe, E. E. Thermal niches of planktonic foraminifera are static throughout glacial-interglacial climate change. *Proc. Natl Acad. Sci. USA* **118**, e2017105118 (2021).

112. 112.

Lonhart, S. I., Jeppesen, R., Beas-Luna, R., Crooks, J. A. & Lorda, J. Shifts in the distribution and abundance of coastal marine species along the eastern Pacific Ocean during marine heatwaves from 2013 to 2018. *Mar. Biodivers. Rec.* **12**, 13 (2019).

113. 113.

Pinsky, M. L., Selden, R. L. & Kitchel, Z. J. Climate-driven shifts in marine species ranges: scaling from organisms to communities. *Ann. Rev. Mar. Sci.* **12**, 153–179 (2020).

114. 114.

Thurman, L. L. et al. Persist in place or shift in space? Evaluating the adaptive capacity of species to climate change. *Front. Ecol. Environ.* **18**, 520–528 (2020).

115. 115.

Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W. & Holt, R. D. A framework for community interactions under climate change. *Trends Ecol. Evol.* **25**, 325–331 (2010).

116. 116.

Dutkiewicz, S., Boyd, P. W. & Riebesell, U. Exploring biogeochemical and ecological redundancy in phytoplankton communities in the global ocean. *Glob. Change Biol.* **27**, 1196–1213 (2021).

117. 117.

Bernhardt, J. R. & Leslie, H. M. Resilience to climate change in coastal marine ecosystems. *Ann. Rev. Mar. Sci.* **5**, 371–392 (2013).

118. 118.

Grant, P. R. et al. Evolution caused by extreme events. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 5–8 (2017).

119. 119.

Gonzalez, A. & Loreau, M. The causes and consequences of compensatory dynamics in ecological communities. *Annu. Rev. Ecol. Evol. Syst.* **40**, 393–414 (2009).

120. 120.

Vallina, S. M. & Le Quéré, C. Stability of complex food webs: resilience, resistance and the average interaction strength. *J. Theor. Biol.* **272**, 160–173 (2011).

121. 121.

Neutel, A. M. et al. Reconciling complexity with stability in naturally assembling food webs. *Nature* **449**, 599–602 (2007).

122. 122.

Ives, A. R. & Cardinale, B. J. Food-web interactions govern the resistance of communities after non-random extinctions. *Nature* **429**, 174–177 (2004).

123. 123.

Nagelkerken, I., Goldenber, S. U., Ferreira, C. M., Ullah, H. & Conne, S. D. Trophic pyramids reorganize when food web architecture fails to adjust to ocean change. *Science* **369**, 829–832 (2020).

124. 124.

Carpenter, S. R. et al. Cascading trophic interactions and lake productivity. *Bioscience* **35**, 634–639 (1985).

125. 125.

Bideault, A. et al. Thermal mismatches in biological rates determine trophic control and biomass distribution under warming. *Glob. Change Biol.* **27**, 257–269 (2021).

126. 126.

Dee, L. E., Okamoto, D., Gårdmark, A., Montoya, J. M. & Miller, S. J. Temperature variability alters the stability and thresholds for collapse of interacting species: species interactions facing variability. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190457 (2020).

127. 127.

Adjeroud, M. et al. Recovery of coral assemblages despite acute and recurrent disturbances on a south central Pacific reef. *Sci. Rep.* **8**, 9680 (2018).

128. 128.

Hughes, T. P. et al. Ecological memory modifies the cumulative impact of recurrent climate extremes. *Nat. Clim. Change* **9**, 40–43 (2019).

129. 129.

Boyd, P. W. et al. Biological responses to environmental heterogeneity under future ocean conditions. *Glob. Change Biol.* **22**, 2633–2650 (2016).

130. 130.

Ainsworth, T. D., Hurd, C. L., Gates, R. D. & Boyd, P. W. How do we overcome abrupt degradation of marine ecosystems and meet the challenge of heat waves and climate extremes? *Glob. Change Biol.* **26**, 343–354 (2020).

131. 131.

Pörtner, H.-O. Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. *J. Exp. Biol.* **213**, 881–893 (2010). **Develops the concept of how other stressors can interact with each other in marine ectotherms.**

132. 132.

Deutsch, C., Ferrel, A., Seibel, B., Portner, H.-O. & Huey, R. B. Climate change tightens a metabolic constraint on marine habitats. *Science* **348**, 1132–1135 (2015).

133. 133.

Deutsch, C., Penn, J. L. & Seibel, B. Metabolic trait diversity shapes marine biogeography. *Nature* **585**, 557–562 (2020).

134. 134.

Kroeker, K. J. et al. Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. *Glob. Change Biol.* **19**, 1884–1896 (2013).

135. 135.

Bertolini, C. & Pastres, R. Tolerance landscapes can be used to predict species-specific responses to climate change beyond the marine heatwave concept: using tolerance landscape models for an ecologically meaningful classification of extreme climate events. *Estuar. Coast. Shelf Sci.* **252**, 107284 (2021).

136. 136.

Le Gland, G., Vallina, S. M., Smith, S. L. & Cermeño, P. SPEAD 1.0 —simulating plankton evolution with adaptive dynamics in a two-trait continuous fitness landscape applied to the Sargasso Sea. *Geosci. Model Dev.* **14**, 1949–1985 (2021).

137. 137.

Merico, A., Bruggeman, J. & Wirtz, K. A trait-based approach for downscaling complexity in plankton ecosystem models. *Ecol. Modell.* **220**, 3001–3010 (2009).

138. 138.

Walworth, N. G., Zakem, E. J., Dunne, J. P., Collins, S. & Levine, N. M. Microbial evolutionary strategies in a dynamic ocean. *Proc. Natl Acad. Sci. USA* **117**, 5943–5948 (2020).

139. 139.

Toseland, A. et al. The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nat. Clim. Change* **3**, 979–984 (2013).

140. 140.

Collins, M. et al. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (eds. Pörtner, H.-O. et al.) Ch. 6 (IPCC, 2021).

141. 141.

Hughes, T. P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).

142. 142.

Cheung, W. W. L., Reygondeau, G. & Frölicher, T. L. Large benefits to marine fisheries of meeting the 1.5 °C global warming target. *Science* **354**, 1591–1594 (2016).

143. 143.

Rashid Sumaila, U. et al. Benefits of the Paris Agreement to ocean life, economies, and people. *Sci. Adv.* **5**, eaau3855 (2019).

144. 144.

Tilbrook, B. et al. An enhanced ocean acidification observing network: from people to technology to data synthesis and information exchange. *Front. Mar. Sci.* **6**, 337 (2019).

145. 145.

Claustre, H., Johnson, K. S. & Takeshita, Y. Observing the global ocean with Biogeochemical-Argo. *Ann. Rev. Mar. Sci.* **12**, 23–48 (2020).

146. 146.

Chai, F. et al. Monitoring ocean biogeochemistry with autonomous platforms. *Nat. Rev. Earth Environ.* **1**, 315–326 (2020).

147. 147.

Fennel, K. et al. Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Front. Mar. Sci.* **6**, 89 (2019).

148. 148.

Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).

149. 149.

Sen Gupta, A. et al. Drivers and impacts of the most extreme marine heatwaves events. *Sci. Rep.* **10**, 19359 (2020).

150. 150.

Holbrook, N. J. et al. Keeping pace with marine heatwaves. *Nat. Rev. Earth Environ.* **1**, 482–493 (2020).

151. 151.

Boyd, P. & Hutchins, D. Understanding the responses of ocean biota to a complex matrix of cumulative anthropogenic change. *Mar. Ecol. Prog. Ser.* **470**, 125–135 (2012).

152. 152.

Thomas, M. K. et al. Temperature–nutrient interactions exacerbate sensitivity to warming in phytoplankton. *Glob. Change Biol.* **23**, 3269–3280 (2017).

153. 153.

Clark, J. R., Daines, S. J., Lenton, T. M., Watson, A. J. & Williams, H. T. P. Individual-based modelling of adaptation in marine microbial populations using genetically defined physiological parameters. *Ecol. Modell.* **222**, 3823–3837 (2011).

154. 154.

Bruggeman, J. & Kooijman, S. A. L. M. A biodiversity-inspired approach to aquatic ecosystem modeling. *Limnol. Oceanogr.* **52**, 1533–1544 (2007).

155. 155.

Small-Lorenz, S. L., Culp, L. A., Ryder, T. B., Will, T. C. & Marra, P. P. A blind spot in climate change vulnerability assessments. *Nat. Clim. Change* **3**, 91–93 (2013).

156. 156.

Sala, E. et al. Protecting the global ocean for biodiversity, food and climate. *Nature* **592**, 397–402 (2021).

157. 157.

Halpern, B. S. et al. An index to assess the health and benefits of the global ocean. *Nature* **488**, 615–620 (2012).

158. 158.

Suryan, R. M. et al. Ecosystem response persists after a prolonged marine heatwave. *Sci. Rep.* **11**, 6235 (2021).

159. 159.

Costanza, R. et al. Changes in the global value of ecosystem services. *Glob. Environ. Change* **26**, 152–158 (2014).

160. 160.

Liquete, C. et al. Current status and future prospects for the assessment of marine and coastal ecosystem services: a systematic review. *PLoS ONE* **8**, e67737 (2013).

161. 161.

Glynn, P. W. & D'Croz, L. Experimental evidence for high temperature stress as the cause of El Niño-coincident coral mortality. *Coral Reefs* **8**, 181–191 (1990).

162. 162.

Eakin, C. M. et al. Caribbean corals in crisis: record thermal stress, bleaching, and mortality in 2005. *PLoS ONE* **5**, e13969 (2010).

163. 163.

Gardner, J., Manno, C., Bakker, D. C. E., Peck, V. L. & Tarling, G. A. Southern Ocean pteropods at risk from ocean warming and acidification. *Mar. Biol.* **165**, 8 (2018).

## Acknowledgements

N.G., M.V. and T.L.F. acknowledge support from the Swiss National Science Foundation (project XEBUS and PP00P2-198897), and from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820989 (project COMFORT). The work reflects only the authors' views; the European Commission and their executive agency are not responsible for any use that may be made of the information the work contains. P.W.B. was primarily supported by the Australian Research Council through a Laureate (FL160100131). P.W.B. also acknowledges support from NSF to the SCOR infrastructure project COBS (Changing Ocean Biological Systems). N.G. thanks F. Desmet, E. Köhn and J. Perreten for evaluating and analysing the regional ocean biogeochemical model simulations for the northeast Pacific Blob event. The GFDL ESM2M simulations were conducted by Friedrich Burger on the Swiss National

Supercomputing Centre. We thank K. Fennel and A. Sen Gupta for their constructive reviews and inputs.

## Author information

### Affiliations

1. Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland

Nicolas Gruber & Meike Vogt

2. Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia

Philip W. Boyd

3. Climate and Environmental Physics, University of Bern, Bern, Switzerland

Thomas L. Frölicher

4. Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

Thomas L. Frölicher

### Contributions

N.G. conceived this Perspective, visualized and wrote it with strong conceptual and editorial input by all of the co-authors. P.W.B. and M.V. led the writing of the biological impact section and T.F. led the analyses of the global model results.

### Corresponding author

Correspondence to [Nicolas Gruber](#).

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Peer review information** *Nature* thanks Katja Fennel and Alex Sen Gupta for their contribution to the peer review of this work.

## Supplementary information

### Supplementary Information

Supplementary Notes and Supplementary Figs. 1–6.

## Rights and permissions

### Reprints and Permissions

## About this article

## Cite this article

Gruber, N., Boyd, P.W., Frölicher, T.L. *et al.* Biogeochemical extremes and compound events in the ocean. *Nature* **600**, 395–407 (2021).  
<https://doi.org/10.1038/s41586-021-03981-7>

- Received: 23 July 2020
- Accepted: 01 September 2021

- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-03981-7>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03981-7>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 15 December 2021](#)

# Sublimation-driven convection in Sputnik Planitia on Pluto

- [Adrien Morison](#) ORCID: [orcid.org/0000-0002-8261-9143<sup>1</sup>](https://orcid.org/0000-0002-8261-9143),
- [Stéphane Labrosse](#) ORCID: [orcid.org/0000-0002-7620-4363<sup>2</sup>](https://orcid.org/0000-0002-7620-4363) &
- [Gaël Choblet](#) ORCID: [orcid.org/0000-0002-0220-9233<sup>3</sup>](https://orcid.org/0000-0002-0220-9233)

*Nature* volume **600**, pages 419–423 (2021)

- 685 Accesses
- 128 Altmetric
- [Metrics details](#)

## Subjects

- [Asteroids, comets and Kuiper belt](#)
- [Fluid dynamics](#)
- [Geodynamics](#)

## Abstract

Sputnik Planitia is a nitrogen-ice-filled basin on Pluto<sup>1</sup>. Its polygonal surface patterns<sup>2</sup> have been previously explained as a result of solid-state convection with either an imposed heat flow<sup>3</sup> or a temperature difference within the 10-km-thick ice layer<sup>4</sup>. Neither explanation is satisfactory, because they do not exhibit surface topography with the observed pattern: flat polygons delimited by narrow troughs<sup>5</sup>. Internal heating produces the

observed patterns<sup>6</sup>, but the heating source in such a setup remains enigmatic. Here we report the results of modelling the effects of sublimation at the surface. We find that sublimation-driven convection readily produces the observed polygonal structures if we assume a smaller heat flux ( $\sim 0.3 \text{ mW m}^{-2}$ ) at the base of the ice layer than the commonly accepted value of  $2\text{--}3 \text{ mW m}^{-2}$  (ref. <sup>7</sup>). Sustaining this regime with the latter value is also possible, but would require a stronger viscosity contrast ( $\sim 3,000$ ) than the nominal value ( $\sim 100$ ) considered in this study.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

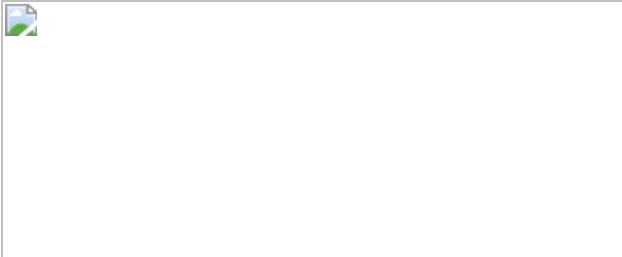
[Rent or Buy](#)

All prices are NET prices.

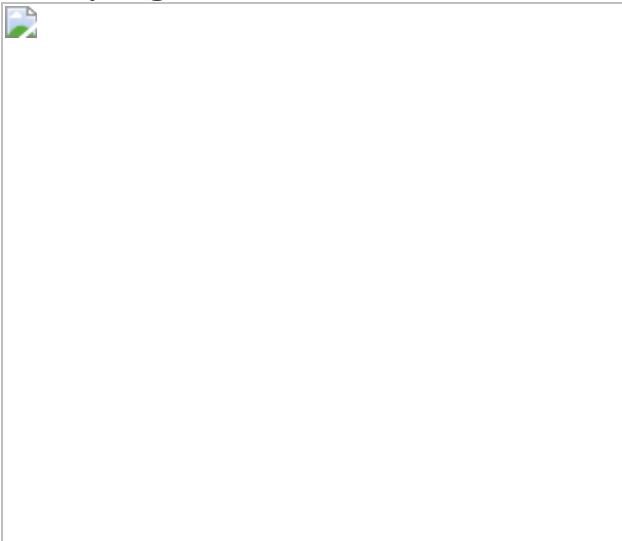
## Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Surface topography as a function of the driving buoyancy source.**



**Fig. 2: Pattern of convection obtained as a function of viscosity contrast and Rayleigh number.**



**Fig. 3: Application of cases in the polygonal regime to SP.**



## Data availability

All results necessary to produce the figures are available in the Zenodo repository <https://doi.org/10.5281/zenodo.5511744>.

## Code availability

The convection code StagYY is the property of Paul J. Tackley and Eidgenössische Technische Hochschule Zürich. Researchers interested in using StagYY should contact Paul J. Tackley (paul.tackley@erdw.ethz.ch). Scripts for the treatment of results and drawing of figures have been written in Python (<https://www.python.org/>) using the StagPy package (<https://github.com/StagPython/StagPy>) and are available in the Zenodo repository, along with the data.

## References

1. 1.

Moore, J. M. et al. The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).

2. 2.

White, O. L. et al. Geological mapping of Sputnik Planitia on Pluto. *Icarus* **287**, 261–286 (2017).

3. 3.

Trowbridge, A. J., Melosh, H. J., Steckloff, J. K. & Freed, A. M. Vigorous convection as the explanation for Pluto's polygonal terrain. *Nature* **534**, 79–81 (2016).

4. 4.

McKinnon, W. B. et al. Convection in a volatile nitrogen-ice-rich layer drives Pluto's geological vigour. *Nature* **534**, 82–85 (2016).

5. 5.

Schenk, P. et al. High-resolution topography of Pluto and Charon: getting down to details. In *Proc. 49th Lunar and Planetary Science Conference* (2018).

6. 6.

Vilella, K. & Deschamps, F. Thermal convection as a possible mechanism for the origin of polygonal structures on Pluto's surface. *J. Geophys. Res.* **122**, 1056–1076 (2017).

7. 7.

Hammond, N. P., Barr, A. C. & Parmentier, E. M. Recent tectonic activity on Pluto driven by phase changes in the ice shell. *Geophys. Res. Lett.* **43**, 6775–6782 (2016).

8. 8.

Mangold, N. Ice sublimation as a geomorphic process: a planetary perspective. *Geomorphology* **126**, 1–17 (2011).

9. 9.

Lliboutry, L. The origin of penitents. *J. Glaciol.* **2**, 331–338 (1954).

10. 10.

Claudin, P., Jarry, H., Vignoles, G., Plapp, M. & Andreotti, B. Physical processes causing the formation of penitentes. *Phys. Rev. E* **92**, 033015 (2015).

11. 11.

Moore, J. M. et al. Sublimation as a landform-shaping process on Pluto. *Icarus* **287**, 320–333 (2016).

12. 12.

Forget, F. et al. A post-new horizons global climate model of Pluto including the N<sub>2</sub>, CH<sub>4</sub> and CO cycles. *Icarus* **287**, 54–71 (2017).

13. 13.

Bertrand, T. et al. The nitrogen cycles on Pluto over seasonal and astronomical timescales. *Icarus* **309**, 277–296 (2018).

14. 14.

Buhler, P. B. & Ingersoll, A. P. Sublimation pit distribution indicates convection cell surface velocities of ~10 cm per year in Sputnik Planitia, Pluto. *Icarus* **300**, 327–340 (2018).

15. 15.

Doumenc, F., Boeck, T., Guerrier, B. & Rossi, M. Transient Rayleigh–Bénard–Marangoni convection due to evaporation: a linear non-normal stability analysis. *J. Fluid Mech.* **648**, 521–539 (2010).

16. 16.

Tackley, P. J. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Phys. Earth Planet. Inter.* **171**, 7–18 (2008).

17. 17.

Davaille, A. & Jaupart, C. Transient high-Rayleigh-number thermal convection with large viscosity variations. *J. Fluid Mech.* **253**, 141–166 (1993).

18. 18.

Solomatov, V. S. Scaling of temperature- and stress-dependent viscosity convection. *Phys. Fluids* **7**, 266–274 (1995).

19. 19.

Nimmo, F. et al. Reorientation of Sputnik Planitia implies a subsurface ocean on Pluto. *Nature* **540**, 94–96 (2016).

20. 20.

Keane, J. T., Matsuyama, I., Kamata, S. & Steckloff, J. K. Reorientation and faulting of Pluto due to volatile loading within Sputnik Planitia. *Nature* **540**, 90–93 (2016).

21. 21.

Eluszakiewicz, J. & Stevenson, D. J. Rheology of solid methane and nitrogen: applications to Triton. *Geophys. Res. Lett.* **17**, 1753–1756 (1990).

22. 22.

Yamashita, Y., Kato, M. & Arakawa, M. Experimental study on the rheological properties of polycrystalline solid nitrogen and methane: implications for tectonic processes on Triton. *Icarus* **207**, 972–977 (2010).

23. 23.

Bertrand, T. et al. Pluto's beating heart regulates the atmospheric circulation: results from high-resolution and multiyear numerical climate simulations. *J. Geophys. Res. Planets* **125**, e2019JE006120 (2020).

24. 24.

Scipioni, F. et al. Pluto's Sputnik Planitia: composition of geological units from infrared spectroscopy. *Icarus* **359**, 114303 (2021).

25. 25.

Earle, A. M. et al. Long-term surface temperature modeling of Pluto. *Icarus* **287**, 37–46 (2017).

26. 26.

Gladstone, G. R. et al. The atmosphere of Pluto as observed by New Horizons. *Science* **351**, aad8866 (2016).

27. 27.

Merlin, F., Lellouch, E., Quirico, E. & Schmitt, B. Triton’s surface ices: distribution, temperature and mixing state from VLT/SINFONI observations. *Icarus* **314**, 274–293 (2018).

28. 28.

Sori, M. M., Bapst, J., Bramson, A. M., Byrne, S. & Landis, M. E. A Wunda-full world? carbon dioxide ice deposits on Umbriel and other Uranian moons. *Icarus* **290**, 1–13 (2017).

29. 29.

Hofgartner, J. D., Buratti, B. J., Hayne, P. O. & Young, L. A. Ongoing resurfacing of KBO Eris by volatile transport in local, collisional, sublimation atmosphere regime. *Icarus* **334**, 52–61 (2019).

30. 30.

Ortiz, J. et al. Albedo and atmospheric constraints of dwarf planet Makemake from a stellar occultation. *Nature* **491**, 566–569 (2012).

31. 31.

Scott, T. Solid and liquid nitrogen. *Phys. Rep.* **27**, 89–157 (1976).

32. 32.

Umurhan, O. et al. Modeling glacial flow on and onto Pluto’s Sputnik Planitia. *Icarus* **287**, 301–319 (2017).

33. 33.

Trouette, B., Chénier, E., Delcarte, C. & Guerrier, B. Numerical study of convection induced by evaporation in cylindrical geometry. *Eur. Phys. J. Spec. Top.* **192**, 83–93 (2011).

34. 34.

Crank, J. *Free and Moving Boundary Problems* (Oxford Univ. Press, 1984).

35. 35.

Hansen, C. J. & Paige, D. A. Seasonal nitrogen cycles on Pluto. *Icarus* **120**, 247–265 (1996).

36. 36.

Ricard, Y., Labrosse, S. & Dubuffet, F. Lifting the cover of the cauldron: convection in hot planets. *Geochem. Geophys. Geosyst.* **15**, 4617–4630 (2014).

37. 37.

Labrosse, S. Hotspots, mantle plumes and core heat loss. *Earth Planet. Sci. Lett.* **199**, 147–156 (2002).

38. 38.

Choblet, G. & Sotin, C. 3D thermal convection with variable viscosity: can transient cooling be described by a steady state scaling law? *Phys. Earth Plan. Inter.* **119**, 321–336 (2000).

39. 39.

McKinnon, W. B. et al. Origin of the Pluto–Charon system: constraints from the New Horizons flyby. *Icarus* **287**, 2–11 (2017).

40. 40.

Néri, A., Guyot, F., Reynard, B. & Sotin, C. A carbonaceous chondrite and cometary origin for icy moons of Jupiter and Saturn. *Earth Planet. Sci. Lett.* **530**, 115920 (2020).

41. 41.

McKinnon, W. et al. Pluto's heat flow: a mystery wrapped in an ocean inside an ice shell. In *Proc. Lunar and Planetary Science Conference* 2715 (2018).

42. 42.

Davidsson, B. et al. The primordial nucleus of comet 67P/Churyumov-Gerasimenko. *Astron. Astrophys.* **592**, A63 (2016).

43. 43.

Bardyn, A. et al. Carbon-rich dust in comet 67P/Churyumov-Gerasimenko measured by COSIMA/Rosetta. *Mon. Not. R. Astron. Soc.* **469**, S712–S722 (2017).

44. 44.

Robuchon, G. & Nimmo, F. Thermal evolution of Pluto and implications for surface tectonics and a subsurface ocean. *Icarus* **216**, 426–439 (2011).

45. 45.

Kamata, S. et al. Pluto's ocean is capped and insulated by gas hydrates. *Nat. Geosci.* **12**, 407–410 (2019).

## Acknowledgements

We are grateful to P. Tackley for providing his mantle convection code StagYY, which we used for our calculations. We gratefully acknowledge support from the Pôle Scientifique de Modélisation Numérique of the ENS de Lyon for the computing resources. G.C. acknowledges the support from

the Agence Nationale de la Recherche (France) under grant number ANR-20-CE49-0010.

## Author information

### Affiliations

1. Physics and Astronomy, University of Exeter, Exeter, UK

Adrien Morison

2. LGL-TPE, Université de Lyon, ENSL, Univ Lyon-1, CNRS, Lyon, France

Stéphane Labrosse

3. Laboratoire de Planétologie et Géodynamique, UMR 6112, Nantes Université, CNRS, Université d'Angers, Nantes, France

Gaël Choblet

### Contributions

All authors developed the theory and designed the study together. A.M. modified the StagYY code; A.M. and S.L. ran the calculations and developed the analysis Python scripts. All authors contributed to the paper.

### Corresponding author

Correspondence to [Stéphane Labrosse](#).

### Ethics declarations

### Competing interests

The authors declare no competing interests.

# Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Topography slope.

Slope of topography for cases # 14 (left) and 15 (right), with the mean value written in the upper left corner.

### Extended Data Fig. 2 Time evolution of output diagnostics for case # 14.

Evolution with time of the surface heat flux (**a**), the mean temperature (**b**), the RMS velocity (**c**), the effective Rayleigh number (**d**) and viscosity contrast (**e**) in the case of Fig. 2b, e of the main text. The first three quantities have been rendered dimensional using the nominal values of the parameters, see Table 1 of the main text.

### Extended Data Fig. 3 Effective temperature as function of effective Rayleigh number for case # 14.

Evolution of the dimensionless temperature as function of the internal heating Rayleigh number, using secular cooling as effective internal heating source (see methods), for case # 14. The dashed line shows a guide to the eye for the expected scaling law in steadily cooling situation. The circled numbers correspond to snapshots, some being presented on Extended Data Fig. 4, the red ones being in the polygonal regime, which is first established at snapshot 18.

### Extended Data Fig. 4 Pattern maturation for case # 14.

Four snapshots (# 2, 10, 18 and 30) of the surface topography (left) and mid-depth temperature anomaly (right) for case # 14, each associated with a circled number on Extended Data Fig. 3.

### Extended Data Fig. 5 Temperature profiles during decaying convection.

Evolution of the horizontal minimum (blue), mean (red) and maximum (green) temperature profiles following the application of a zero heat flux at the upper surface of an established convecting system, case #14.

### Extended Data Fig. 6 Decaying time-scale of convection.

Time evolution of the temperature anomalies (**a**, **b**) and the RMS velocity (**c**) after the surface heat flow has been set to zero. (**a**) shows the vertical mean of  $T_{max}(z) - T_{min}(z)$ . (**b**) shows the RMS of  $\langle T - \bar{T} \rangle(z)$ , with  $\langle \bar{T} \rangle(z)$  the horizontal mean of temperature. The exponential fit has been performed on values for time  $\geq 55$  kyr.

### Extended Data Fig. 7 Time evolution of output diagnostics for a time-varying boundary condition.

Time evolution of the same diagnostics as in Extended Data Fig. 2 for a time varying boundary condition at the surface. The insets in each panel shows a zoom on the last 5 periods of the oscillations.

### Extended Data Fig. 8 Convection pattern for a time-varying boundary condition.

Surface topography and mid-depth temperature anomaly at the end of the time varying run presented on Extended Data Fig. 7.

### Extended Data Fig. 9 Evolution of the layer with an imposed bottom heat flux $q_{bot} = 0.26 \text{ mW m}^{-2}$ .

Panel **(a)** shows the evolution of the surface and bottom heat fluxes, **(b)** the mean temperature, **(c)** the RMS velocity, **(d)** the effective Rayleigh number, **(e)** the effective viscosity contrast and the bottom ones, snapshots of the topography at times written on their respective upper left corners.

### Extended Data Fig. 10 Evolution of the layer with an imposed bottom heat flux $q_{bot} = 1.3 \text{ mW m}^{-2}$ .

Panel **(a)** shows the evolution of the surface and bottom heat fluxes, **(b)** the mean temperature, **(c)** the RMS velocity, **(d)** the effective Rayleigh number, **(e)** the effective viscosity contrast and the bottom ones, snapshots of the topography at times written on their respective upper left corners.

### Extended Data Fig. 11 Evolution of the layer for $Ra_\infty = 10^9$ and $\{\{\boldsymbol{R}\}\}_{\{\{\boldsymbol{\eta}\}\}}^{\{\infty\}} = \{\{\bf{10}\}\}_{\{\{\bf{4}\}\}}^{\{\infty\}}$ with a zero imposed bottom heat flux.

Panel **(a)** shows the evolution of the surface and bottom heat fluxes, **(b)** the mean temperature, **(c)** the RMS velocity, **(d)** the effective Rayleigh number, **(e)** the effective viscosity contrast and the bottom ones, snapshots of the topography at times written on their respective upper left corners.

### Extended Data Fig. 12 Evolution of the layer for $Ra_\infty = 10^9$ and $\{\{\boldsymbol{R}\}\}_{\{\{\boldsymbol{\eta}\}\}}^{\{\infty\}} = \{\{\bf{10}\}\}_{\{\{\bf{4}\}\}}^{\{\infty\}}$ with an imposed bottom heat flux $q_{bot} = 1.3 \text{ mW m}^{-2}$ .

Panel **(a)** shows the evolution of the surface and bottom heat fluxes, **(b)** the mean temperature, **(c)** the RMS velocity, **(d)** the effective Rayleigh number, **(e)** the effective viscosity contrast and the bottom ones, snapshots of the topography at times written on their respective upper left corners.

### **Extended Data Table 1 Input parameters and output diagnostics**

## **Supplementary information**

### **41586\_2021\_4095\_MOESM1\_ESM.pdf**

Supplementary information This file presents some diagnostics for the last snapshot of each studied case, as listed in Extended Data Table 1. For each case, the second column shows the surface topography, the temperature anomaly at mid-depth and the vertical profile of the r.m.s. of the horizontal velocity, all figures being dimensionless.

### **Peer Review File**

### **Supplementary Video 1**

Supplementary video showing the evolution with time of the topography (left) and mid-depth temperature anomaly in case 14.1.

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Morison, A., Labrosse, S. & Choblet, G. Sublimation-driven convection in Sputnik Planitia on Pluto. *Nature* **600**, 419–423 (2021).  
<https://doi.org/10.1038/s41586-021-04095-w>

- Received: 03 April 2021
- Accepted: 04 October 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021

- DOI: <https://doi.org/10.1038/s41586-021-04095-w>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## [Pluto's strange ice patterns explained by new theory](#)

- Nick Petrić Howe
- Benjamin Thompson

Nature Podcast 15 Dec 2021

## [Pluto's mysterious polygons explained](#)

- Shamina Bundell

Nature Video 17 Dec 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04095-w>

- Article
- Open Access
- [Published: 15 December 2021](#)

# Direct limits for scalar field dark matter from a gravitational-wave detector

- [Sander M. Vermeulen](#) [ORCID: orcid.org/0000-0003-4227-8214](#)<sup>1</sup>,
- [Philip Relton](#)<sup>1</sup>,
- [Hartmut Grote](#) [ORCID: orcid.org/0000-0002-0797-3943](#)<sup>1</sup>,
- [Vivien Raymond](#)<sup>1</sup>,
- [Christoph Affeldt](#)<sup>2</sup>,
- [Fabio Bergamin](#)<sup>2</sup>,
- [Aparna Bisht](#)<sup>2</sup>,
- [Marc Brinkmann](#)<sup>2</sup>,
- [Karsten Danzmann](#)<sup>2</sup>,
- [Suresh Doravari](#)<sup>2</sup>,
- [Volker Kringel](#)<sup>2</sup>,
- [James Lough](#) [ORCID: orcid.org/0000-0002-5160-0239](#)<sup>2</sup>,
- [Harald Lück](#)<sup>2</sup>,
- [Moritz Mehmet](#)<sup>2</sup>,
- [Nikhil Mukund](#) [ORCID: orcid.org/0000-0002-8666-9156](#)<sup>2</sup>,
- [Séverin Nadjji](#)<sup>2</sup>,
- [Emil Schreiber](#)<sup>2</sup>,
- [Borja Sorazu](#)<sup>3</sup>,
- [Kenneth A. Strain](#)<sup>2,3</sup>,
- [Henning Vahlbruch](#) [ORCID: orcid.org/0000-0003-2357-2338](#)<sup>2</sup>,
- [Michael Weinert](#)<sup>2</sup>,
- [Benno Willke](#)<sup>2</sup> &
- [Holger Wittel](#)<sup>2</sup>

*Nature* volume 600, pages 424–428 (2021)

- 446 Accesses
- 130 Altmetric

- [Metrics details](#)

## Subjects

- [Dark energy and dark matter](#)
- [Quantum metrology](#)
- [Quantum optics](#)

## Abstract

The nature of dark matter remains unknown to date, although several candidate particles are being considered in a dynamically changing research landscape<sup>1</sup>. Scalar field dark matter is a prominent option that is being explored with precision instruments, such as atomic clocks and optical cavities<sup>2,3,4,5,6,7,8</sup>. Here we describe a direct search for scalar field dark matter using a gravitational-wave detector, which operates beyond the quantum shot-noise limit. We set new upper limits on the coupling constants of scalar field dark matter as a function of its mass, by excluding the presence of signals that would be produced through the direct coupling of this dark matter to the beam splitter of the GEO600 interferometer. These constraints improve on bounds from previous direct searches by more than six orders of magnitude and are, in some cases, more stringent than limits obtained in tests of the equivalence principle by up to four orders of magnitude. Our work demonstrates that scalar field dark matter can be investigated or constrained with direct searches using gravitational-wave detectors and highlights the potential of quantum-enhanced interferometry for dark matter detection.

[Download PDF](#)

## Main

Laser interferometers have very high sensitivity to minute length changes and have facilitated many gravitational-wave detections over the past few years<sup>9,10</sup>. In addition to their merit in astrophysics, the detection of gravitational waves has also shed light on fundamental physics questions and several links may exist between gravitational waves and dark matter<sup>11</sup>. Owing to their excellent sensitivity at or beyond quantum limits, gravitational-wave detectors (or precision interferometers of a similar type) can be used directly for fundamental physics, without the mediation of gravitational waves. Examples include a possible search for vacuum birefringence<sup>12</sup> and the search for signatures of quantum gravity<sup>13,14,15</sup>. Several ideas have been put forward as to how different candidates of dark matter can directly couple to gravitational-wave detectors, ranging from scalar field dark matter<sup>4,16</sup> to dark photon dark matter<sup>17</sup> and to

clumpy dark matter coupling gravitationally or through an additional Yukawa force<sup>18</sup>. Upper limits on dark photon dark matter have already been set in a small mass band using data from the first observational run (O1) of the Advanced LIGO gravitational-wave detectors<sup>19</sup>.

In this work, we conduct a direct search for scalar field dark matter using a gravitational-wave detector, the quantum-enhanced GEO600 interferometer, and set new upper limits on the parameters of such dark matter.

## Theory

Models of weakly coupled low-mass ( $\ll 1$  eV) scalar fields predict that such particles could be produced in the early Universe through a vacuum misalignment mechanism and would manifest as a coherently oscillating field<sup>2,4</sup>,

$$\$ \$ \varphi(t, \{r\}) = \{\varphi_0\} \{c\} \{o\} \{s\} (\{\omega\} \{\varphi_t - \{\bf{k}\} \cdot \dot{\varphi}\} \cdot \bf{v}_\text{obs}), \$ \$$$

(1)

where  $\omega_\phi = (m_\phi c^2)/\hbar$  is the angular Compton frequency and  $\mathbf{k}_\phi = (m_\phi \mathbf{v}_\text{obs})/\hbar$  is the wave vector, with  $m_\phi$  the mass of the field and  $\mathbf{v}_\text{obs}$  the velocity relative to the observer. The amplitude of the field can be set as  $\{\varphi_0\} = \sqrt{2 \{\rho_\text{local}\}} / \{m\}$  under the assumption that this scalar field constitutes the local dark matter density  $\rho_\text{local}$  (ref. [20](#)).

Moreover, these models predict that such dark matter would be trapped and virialized in gravitational potentials, leading to a Maxwell–Boltzmann-like distribution of velocities  $\mathbf{v}_\text{obs}$  relative to an observer. As non-zero velocities produce a Doppler shift of the observed dark matter field frequency, this virialization results in the dark matter field having a finite coherence time or, equivalently, a spread in observed frequency (linewidth)<sup>[17,21](#)</sup>. The linewidth is determined by the virial velocity, which is given by the depth of the gravitational potential. For dark matter trapped in the galactic gravity potential, as in the standard galactic dark matter halo model, the expected linewidth is  $\Delta\omega_\text{obs}/\omega_\text{obs} \sim 10^{-6}$ . Certain kinds of scalar particles, such as relaxion dark matter <sup>[22,23](#)</sup>, may also form gravitationally bound objects and be captured in the gravitational potential of the Earth or the Sun, producing a local dark matter overdensity where the field has a much narrower linewidth<sup>[24](#)</sup>. The observed dark matter frequency is further modulated by the motion of the Earth with respect to the centre of mass of the local dark matter.

Scalar field dark matter could couple to the fields of the standard model (SM) in numerous ways. Such a coupling, sometimes called a ‘portal’, is modelled by the addition of a parameterized interaction term to the SM Lagrangian<sup>25,26</sup>. In this paper, we consider linear interaction terms with the electron rest mass  $m_e$  and the electromagnetic field tensor  $F_{\mu\nu}$ :

$$\text{L} \supset \frac{\varphi}{\Lambda} \gamma_\mu F_{\mu\nu} F^{\mu\nu} - \frac{\varphi}{\Lambda} \frac{e}{m_e} \bar{\psi} \psi, \quad (2)$$

where  $\psi$  and  $\bar{\psi}$  are the SM electron field and its Dirac conjugate, respectively, and  $\Lambda_\gamma$  and  $\Lambda_e$  parameterize the coupling. Specific types of scalar dark matter, such as the hypothetical moduli and dilaton fields motivated by string theory, have couplings to the quantum chromodynamics part of the SM as well<sup>27,28,29</sup>.

The addition of the terms in equation (2) to the SM Lagrangian entails changes in the fine structure constant  $\alpha$  and the electron rest mass  $m_e$  (refs. 2,3). The apparent variation in these fundamental constants, in turn, changes the lattice spacing and electronic modes of a solid, driving changes in its size  $l$  and refractive index  $n$ :

$$\frac{\delta l}{l} = -\left(\frac{\delta \alpha}{\alpha} + \frac{\delta m_e}{m_e}\right), \quad (3)$$

$$\frac{\delta n}{n} = -5 \times 10^{-3} \left(2 \frac{\delta \alpha}{\alpha} + \frac{\delta m_e}{m_e}\right), \quad (4)$$

where  $\delta x$  denotes a change in the parameter  $x$ :  $x \rightarrow x + \delta x$ . Equations (3) and (4) hold in the adiabatic limit, which applies for a solid with a mechanical resonance frequency much higher than  $\omega_\phi$  (the driving frequency)<sup>16,28,30</sup>.

Laser interferometers for gravitational-wave detection are modified Michelson interferometers with very high sensitivity to differential changes in the optical path length of their arms. The thin cylindrical beam splitter in such an instrument interacts asymmetrically with light from the two arms, as the front surface has a 50% reflectivity and the back surface has an anti-reflective coating. Therefore, a change in the size ( $\delta l$ ) and index of refraction ( $\delta n$ ) of the beam splitter affects the two arms differently and produces an effective difference in the optical path lengths of the arms  $L_{x,y}$

$$\begin{aligned} & \delta(L_x - L_y) \approx \sqrt{2} \left[ \frac{1}{n} - \frac{1}{2} \right] \delta l + \delta n \end{aligned} \quad (5)$$

This expression includes a correction to equation (17) in ref. [16](#). In addition, a geometrical correction ( $\approx 6.4\%$ ) from Snell's law is applied to equations [\(5\)](#) and [\(6\)](#) for calculating the results below.

The mirrors in the arms of gravitational-wave interferometers would also undergo changes in their size and index of refraction, but as the wavelength of the dark matter field is much greater than the distance between the arm mirrors ( $\lambda_\phi/L \gtrsim 10^3$ ) for all frequencies of interest here, and because the mirrors have roughly the same thickness, the effect is almost equal in both arms and therefore does not produce a dominant signal.

The interferometer most sensitive to potential dark matter signals is the GEO600 detector, as it has the highest sensitivity to optical phase differences between the two arms. The squeezed vacuum states of light currently used in this instrument allow for a world-record quantum noise reduction of 6 dB (ref. [31](#)). Although other gravitational-wave detectors (LIGO/Virgo) are more sensitive to gravitational waves through the use of longer arms, including Fabry–Pérot cavities, these do not boost their sensitivity to signals induced at the beam splitter, so their relative sensitivity to scalar dark matter is lower[16](#).

From equations [\(1\)](#)–[\(5\)](#), it follows that an oscillating scalar dark matter field is expected to produce a Doppler-shifted and Doppler-broadened signal in the GEO600 interferometer of the form

$$\begin{aligned} & \delta(L_x - L_y) \approx \left( \frac{1}{\Lambda_{\text{rm}} \gamma} + \frac{1}{\Lambda_{\text{rm}} e} \right) \left( \frac{n}{\hbar} \right) \sqrt{2 \rho_{\text{local}}} m \varphi \cos(\omega_{\text{obs}} t) \end{aligned} \quad (6)$$

where we have neglected the contribution of the refractive index changes to the signal, as it is three orders of magnitude smaller than that of the size changes. Given this prediction, we can examine the data from the interferometer for the presence of such signals and, if none are found, place upper limits on the mass and coupling constants of scalar field dark matter.

**Fig. 1: A typical amplitude spectrum produced with frequency bins that are tuned to the expected dark matter linewidth using the modified LPSD technique.**

---

 **figure1**

---

The black line indicates the amplitude spectrum. The noise spectrum was estimated at each frequency bin from neighbouring bins to yield the local noise median (blue) and 95% confidence level (CL, green). Peaks (red) above this confidence level were considered candidates for dark matter signals and analysed further.

## Results

The GEO600 interferometer<sup>[32](#)</sup> has been in joint observing runs with the Advanced LIGO detectors since 2015, primarily to look for gravitational waves. We performed spectral analysis on seven  $T \sim 10^5$  s segments of strain data from the GEO600 interferometer (acquired in 2016 and 2019) using a modified version of the logarithmic power spectral density (LPSD) technique<sup>[33](#)</sup>, which was designed to

produce spectra with logarithmically spaced frequencies. Using this algorithm to perform discrete Fourier transforms (DFTs) with a frequency-dependent length, we created spectra in which each frequency bin was made to have a width equal to the Doppler-broadened linewidth of potential signals from scalar field dark matter in a galactic halo. A typical spectrum created using this approach is shown in Fig. 1. This method yields, in theory, the maximum attainable signal-to-noise ratio (SNR), given a certain amount of data (see Methods)<sup>21,34</sup>. A matched filtering approach is not feasible, as the phase of the signal varies stochastically.

**Fig. 2: Constraints on the coupling parameters  $A_\gamma$  and  $A_e$  as a function of the mass of the field  $m_\phi$  for scalar field dark matter as in the basic scalar scenario.**



Dashed lines represent constraints on the electron coupling  $\Lambda_e$  and dotted lines represent constraints on the photon coupling  $\Lambda_\gamma$ , at the 95% confidence level. The green region denotes the parameter space excluded in the current study through the spectral analysis of data from the GEO600 gravitational-wave detector. Other coloured regions indicate parameter spaces excluded through previous direct experimental searches: Hees et al.<sup>6</sup> (blue), Van Tilburg et al.<sup>5</sup> (yellow), Kennedy et al.<sup>45</sup> (brown), Aharony et al.<sup>38</sup> (magenta), Branca et al.<sup>40</sup> (purple), Savalle et al.<sup>8</sup> (cyan) and Antypas et al.<sup>39</sup> (red). The black curves and grey regions correspond to previous constraints from ‘fifth-force’ searches/tests of the equivalence principle; to our knowledge, the most stringent such constraints for this dark matter scenario are from the MICROSCOPE experiment<sup>7,43</sup> (lower curves at low mass) and the Cu/Pb torsion pendulum experiment performed by the Eöt-Wash group<sup>26,37,41</sup> (at higher masses).

We analysed the amplitude spectra of all seven strain data segments for the presence of dark matter signals by looking for significant peaks in the underlying noise. Peaks were considered candidates when there was a less than 5% probability of the local maximum being due to noise, where we compensated for the look-elsewhere effect using a large trial factor ( $\sim 10^6$ ).

This analysis found approximately  $10^4$  peaks above the 95% confidence level ( $>5.6\sigma$ ), where the total error includes a frequency-dependent amplitude calibration error of up to 30% inherent to GEO600 data<sup>35</sup>. The frequency and amplitude stability of the peaks in time was then evaluated by cross-checking all candidates between spectra. Candidate peaks were rejected if their centre frequencies differed between spectra by more than the Doppler shift expected from the motion of the Earth around the Sun through a galactic dark matter halo<sup>36</sup>. Peaks were also rejected if their amplitude changed significantly ( $\geq 5\sigma$ ) between spectra.

Using this procedure, we eliminated all but 14 candidate peaks, where the vast majority (>99%) of peaks were rejected because they did not appear in all datasets within the centre frequency tolerance.

These 14 candidate peaks were subjected to further analysis to investigate whether their properties matched that of a dark matter signal. Thirteen of the peaks were found to have insufficient width to be caused by dark matter ( $\Delta f_{\text{peak}}/\Delta f_{\text{DM}} \lesssim 10$ , see Methods). The remaining candidate peak had sufficient frequency spread to be due to dark matter, but additional analysis showed that this signal has a coherence time much greater than that expected for a galactic halo dark matter signal of that frequency ( $\tau_c^{\text{peak}}/\tau_c^{\text{GH}} > 10$ , see Methods). This leaves open the possibility of the signal being due to scalar dark matter gravitationally bound to Earth, such as in a relaxion halo. However, additional investigations revealed that this signal was not present in data

acquired with independent electronics, whereas the noise and the other signals from the interferometer were. The peak was, therefore, rejected and is suspected to be an artefact from a timing signal in the main data acquisition electronics (see Methods for details).

**Fig. 3: Constraints on the coupling parameters  $\Lambda_\gamma$  and  $\Lambda_e$  as a function of the mass of the field  $m_\phi$  for scalar field dark matter as in the dilaton/modulus scenario and the relaxion halo scenario.**

 figure3

The dilaton/modulus scenario is shown on the left and the relaxion halo scenario is shown on the right. Dashed lines represent constraints on the electron coupling  $\Lambda_e$  and dotted lines represent constraints on the photon coupling  $\Lambda_\gamma$ , at the 95% confidence level. The green region denotes the parameter space excluded in the current study through the spectral analysis of data from the GEO600 gravitational-wave detector. Other coloured regions indicate parameter spaces excluded through previous direct experimental searches<sup>5,6,38,39,40,45,46</sup> (see caption of Fig. 2). The black curves and grey regions correspond to previous constraints from ‘fifth-force’ searches/tests of the equivalence principle; to our knowledge, the most stringent such constraints for this dark matter scenario are from the MICROSCOPE experiment<sup>26,43</sup> (lower curves at low mass) and the Be/Ti torsion pendulum experiment performed by the Eöt-Wash group<sup>26,42</sup> (at higher masses). The constraints for the relaxion halo scenario from direct experimental searches have been obtained by rescaling the originally reported constraints to account for the mass-dependent local overdensities as proposed in ref. 24. This produces novel constraints not reported before for relaxion halo dark matter from the results of refs. 5,6,38,40,45. The fifth force/equivalence principle constraints are independent of the local dark matter density and are, thus, unchanged.

Having determined that all major peaks in the amplitude spectrum are not caused by scalar field dark matter, we can set constraints on the parameters of such dark matter at a 95% confidence level (corresponding to  $5.6\sigma$  above the noise floor), using equation (6). We apply our results to three different scalar dark matter scenarios considered in the literature:

Basic scalar (Fig. 2): the scalar field dark matter is assumed to interact with the SM as given by the terms in equation (2) and is further assumed to be homogeneously distributed over the Solar System with a density of  $\rho_{\text{GH}} = 0.4 \text{ GeV cm}^{-3}$ , as in the standard galactic dark matter halo model<sup>20</sup>.

Dilaton/modulus (Fig. 3, left): in addition to the coupling to the electromagnetic sector as in equation (2), the field is assumed to have couplings to the quantum chromodynamics sector and the coupling to the gluon field is assumed to be dominant<sup>27,28,29,37</sup>. The local dark matter density is taken to be  $\langle \rho \rangle_{\{\rm GH\}}$ . Compared with the basic scalar, this scenario is subject to additional limits from tests of the equivalence principle, but is equally constrained by our result and those of other direct searches.

Relaxion halo (Fig. 3, right): in this scenario, the scalar field effectively couples to the SM as in the dilaton/modulus scenario, but these couplings arise through mixing with the Higgs boson<sup>22,23</sup>. The local dark matter density in this scenario is taken to be dominated by a relaxion halo gravitationally bound to Earth, which leads to a local overdensity that depends on the mass of the field and reaches values of up to  $\rho_{\text{local}}/\rho_{\text{GH}} = 10^{11}$  (ref. 24) for the mass range constrained in this work.

For each scenario, we set constraints on the electron and photon coupling parameters  $\Lambda_\gamma$  and  $\Lambda_e$ , respectively, as a function of the mass of the field  $m_\phi$  (where, for each coupling constant, we assume the other to be zero); the constraints are plotted in Figs. 2 and 3, together with previous upper limits. For the relaxion halo scenario, we assumed a mass-dependent halo density, as described in ref. 24.

Constraints from other direct experimental dark matter searches include those from various atomic spectroscopy experiments<sup>5,6,38,39</sup>, a search using an optical cavity<sup>8</sup> and a resonant mass detector<sup>40</sup>. Tests of the equivalence principle using, for example, torsion balances<sup>37,41,42</sup> have also been used to set constraints on the parameters of undiscovered scalar fields; these bounds assume that the scalar field manifests as a ‘fifth force’ and is sourced by a test mass (for example, the Earth)<sup>7,26,43</sup>. The constraints on scalar fields inferred from these experiments generally depend on the composition and topography of the test masses and are independent of the local dark matter density.

## Conclusions

In this paper, we presented a search for signals of scalar field dark matter in the data of a gravitational-wave detector. Scalar field dark matter would cause oscillations of the fine structure constant and electron mass, which—in turn—drive oscillations of the size and index of refraction of the beam splitter in an interferometer. This would therefore produce an oscillatory signal in a gravitational-wave detector at a frequency set by the mass of the dark matter particle.

As very extensive classical noise mitigation is used in gravitational-wave detectors, quantum technologies such as squeezed light can provide a large increase in sensitivity. Such technologies facilitate measurements beyond the shot-noise quantum limit and yield unprecedented sensitivity to scalar field dark matter in a wide mass range.

In addition, by tuning the frequency bin widths to the expected dark matter linewidth, our spectral analysis method improves on the analyses used in previous work that set constraints on dark photons using data from gravitational-wave detectors and other searches for scalar fields in frequency space. In contrast to these other efforts, the spectral analysis presented here yields the optimal SNR for potential dark matter signals across the full frequency range.

We excluded the presence of such signals in the data of the GEO600 gravitational-wave detector, thereby setting new lower limits on dark matter couplings at up to  $\Lambda_\gamma$ ,  $\Lambda_e = 3 \times 10^{19} \text{ GeV}$  for dark matter masses between  $10^{-13}$  and  $10^{-11} \text{ eV}$ . These constraints improve on the current limits in the mass range obtained with atomic spectroscopy experiments by more than six orders of magnitude and are up to four orders of magnitude more stringent than previous bounds from tests of the equivalence principle for some dark matter scenarios.

Tighter constraints on scalar field dark matter in various mass ranges can be set in the future using yet-to-be-built gravitational-wave detectors or other similar precision interferometers. Using the same methods as in this work, these instruments would allow new limits to be set across their characteristic sensitive frequency range. Moreover, by slightly modifying the optics in such interferometers—for example, by using mirrors of different thicknesses in each interferometer arm—their sensitivity to scalar field dark matter could be improved further<sup>16</sup>. Through the reduction of losses, quantum technologies such as squeezed light are also expected to improve, allowing for increasing noise mitigation<sup>44</sup>. These and other future technological advances make precision interferometers operating beyond quantum limits indispensable tools for dark matter detection and fundamental physics in general.

# Methods

## Spectral estimation

Spectral analysis was performed using a modified version of the LPSD technique<sup>33</sup>. This technique is designed to produce spectral estimates with logarithmically spaced frequencies and therefore enables the production of spectral estimates with a frequency-dependent bin width. Using this technique, we subdivided the  $\sim 10^5$  s data segments into

$$\text{N}_f = \left\lfloor \frac{T - \tau_{coh}}{\tau_{coh} \ln(1/\xi)} + 1 \right\rfloor \quad (7)$$

smaller overlapping subsegments  $S_f^k(t)$  with a length equal to the expected coherence time  $\tau_{coh}(f)$ , of a dark matter signal at a frequency  $f$ , where  $\xi \in [0, 1]$  is the fractional overlap of the subsegments and  $k \in [1, N_f]$ . As the expected coherence time and linewidth is frequency dependent, this subdivision is unique for every frequency of interest. After subdivision, the subsegments were multiplied with a Kaiser window function  $W_f(t)$  and subjected to a DFT at a single frequency:

$$a^k(f) = \sum_{t=0}^{T_{DFT}} S_f^k(t) e^{-j2\pi f t}, \quad (8)$$

with  $T_{DFT} = \tau_{coh}(f)$ , where  $a^k(f)$  is the complex spectral estimate at frequency  $f$  for the  $k$ th subsegment. Frequency points are chosen by dividing the interval between the chosen minimum frequency (50 Hz) and the Nyquist frequency ( $\approx 8.2$  kHz) by the dark matter linewidth and then rounding the resulting number of bins to the nearest integer to set the final frequency points and bin widths. The absolute squared magnitudes  $|a^k(f)|^2$  are averaged over the subsegments to obtain the power spectrum

$$P(f) = \frac{C}{N_f} \sum_{k=1}^{N_f} |a^k(f)|^2, \quad (10)$$

where  $C$  is a normalization factor. The spectra used in the analysis were made with a bin width equal to the expected linewidth of galactic dark matter ( $\Delta\omega/\omega \approx 10^{-6}$ , see ref. 21). The amplitude spectrum ( $A(f) = \sqrt{P(f)}$ ) created in this way comprises approximately  $5 \times 10^6$  frequency bins between 50 Hz and 6 kHz.

The SNR for galactic dark matter signals in such a spectrum is optimal given a certain amount of data (see Sec. 'Validation of methods') and can only be further improved by analysing more data. Additional data would allow for more averaging, which decreases the variance of the spectrum as the square root of the amount of data, such that the sensitivity approaches the noise floor. The noise floor can be lowered using longer DFT lengths at the cost of reduced SNR, but this is subject to severely diminishing returns; the sensitivity can only be improved by a factor proportional to the fourth root of the amount of data needed<sup>21</sup> (and the computation time scales with the product of DFT length and the amount of data<sup>33</sup>). Computation times for the spectra used in this work are approximately 10 s per frequency bin for each  $\sim 10^5$  s dataset, or about  $10^4$  CPU hours per spectrum.

## Estimation of noise statistics

The local noise parameters were estimated at every frequency bin from  $w = 5 \times 10^4$  neighbouring bins. This method allows the underlying noise distribution to be estimated in a way that is independent of narrow ( $\ll w$ ) spectral features (such as those due to mechanical excitation of the mirror suspensions), under the assumption that the underlying noise spectrum is locally flat (that is, the autocorrelation length of the noise spectrum is assumed to be  $\gg w$ ). Therefore, the choice of  $w$  represents a trade-off between erroneously assuming instrumental spectral artefacts or signals to be features of the underlying noise spectrum versus erroneously assuming features of the underlying noise spectrum to be instrumental spectral artefacts or signals.

## Follow-up analysis of candidates

As mentioned above, 14 candidate peaks remained after cross-checking spectra taken at different times. Thirteen of these peaks were found to have insufficient width to be dark matter signals. Further investigation of each of these candidates found that shifting the bin centre frequencies by an amount much smaller than the expected linewidth of dark matter signals of that frequency and amplitude and recomputing the spectra did not reproduce the peak. Additional work revealed that these 13 candidate peaks were not present in spectra created using the same data and the same LPSD algorithm implemented in a different programming language, whereas the noise floor and other spectral features were reproduced identically. Therefore, these peaks are probably artefacts of the numerical implementation of the LPSD technique.

The coherence time of the single remaining candidate peak was investigated by evaluating its height in the amplitude spectrum as a function of the DFT length (see 'Validation of methods'). The height of the peak did not decrease for DFT lengths more than an order of magnitude greater than the expected dark matter coherence time, showing a coherence time much greater than that expected for a galactic dark matter

signal of that frequency. To find the origin of the signal and to check whether it could be due to the theoretically more coherent relaxion halo dark matter, we performed spectral analysis on data acquired on an independent data acquisition system. The signal was not present in these data, whereas both noise and other signals from the interferometer were. This fact, in combination with high-resolution ( $\Delta f/f \sim 10^{-7}$ ) spectra revealing that the frequency at which the peak occurs is very close to and indistinguishable from an integer ( $f_{\text{peak}} = 224 \pm (2 \times 10^{-5})$  Hz), implies that the signal is most likely an artefact of a timing signal in the main data acquisition electronics.

## Validation of methods

To validate several aspects of our analysis methods, we simulated dark matter signals and injected them into sets of real and simulated data. The dark matter signals were created by superposing approximately  $10^2$  sinusoids at frequencies linearly spaced around a centre frequency (the simulated Doppler-shifted dark matter Compton frequency), where the amplitude of each sinusoid is given by the quasi-Maxwellian dark matter line shape proposed in ref. [21](#) scaled by a simulated dark matter coupling constant; the relative phases of the sinusoids are randomized to capture the thermalization of the scalar field dark matter.

To test the spectral estimation, signal search and candidate rejection, a blind injection of simulated dark matter signals into several GEO600 datasets was performed, where the frequency, amplitude and number of signals was masked to the authors. All injected signals were recovered at their Compton frequency and at an amplitude corresponding to the hypothetical coupling constant and were subsequently identified through cross-checks between spectra as persistent candidate dark matter signals.

The previously proposed [17,21](#) and herein used condition of setting the frequency bin widths equal to the expected dark matter linewidth for attaining optimal SNR was tested using simulated dark matter signals as well. Mock dark matter signals and monochromatic sine signals were injected into real GEO600 data and Gaussian noise and spectra were made for which the width of the frequency bins  $\Delta f_{\text{bin}}$  (and, correspondingly, the length of the DFTs  $T_{\text{DFT}}$ ) was varied over four orders of magnitude. The recovered amplitude of signals injected into GEO600 data in spectra created using the LPSD algorithm is plotted in Extended Data Fig. [1](#) (left). This shows that the recovered amplitude of signals starts to decrease as the DFT length exceeds the coherence time (a monochromatic sine has infinite coherence time) and validates the rejection of the remaining candidate signal above, as its amplitude was found to be roughly constant for  $T_{\text{DFT}}/\tau_c > 10$ . The recovered SNR of signals injected into Gaussian noise in spectra created using Welch's method [47](#) is plotted in Extended Data Fig. [1](#) (right), which confirms that the SNR is maximal when the frequency bin width

is roughly equal to the full width at half maximum  $\Delta f_{\text{DM}}$  of the spectral line shape of the signal. This is a consequence of the aforementioned decrease in recovered amplitude for smaller bin widths and the scaling of white Gaussian noise.

## Data availability

The upper limit data in Figs. 2 and 3 and intermediate results, such as the spectrum in Fig. 1, are available from the corresponding author on request. The raw data used for the full analysis comprise about 80 GB and are available from the corresponding author on reasonable request.

## Code availability

The code used for this analysis has been released and can be found at <https://github.com/philrelton/Scalar-Dark-Matter-LPSD>.

## References

1. 1.

Bertone, G. & Tait, T. M. P. A new era in the search for dark matter. *Nature* **562**, 51–56 (2018).

2. 2.

Arvanitaki, A., Huang, J. & Van Tilburg, K. Searching for dilaton dark matter with atomic clocks. *Phys. Rev. D* **91**, 015015 (2015).

3. 3.

Derevianko, A. & Pospelov, M. Hunting for topological dark matter with atomic clocks. *Nat. Phys.* **10**, 933–936 (2014).

4. 4.

Stadnik, Y. V. & Flambaum, V. V. Can dark matter induce cosmological evolution of the fundamental constants of nature? *Phys. Rev. Lett.* **115**, 201301 (2015).

5. 5.

Van Tilburg, K., Leefer, N., Bougas, L. & Budker, D. Search for ultralight scalar dark matter with atomic spectroscopy. *Phys. Rev. Lett.* **115**, 011802 (2015).

6. 6.

Hees, A., Guéna, J., Abgrall, M., Bize, S. & Wolf, P. Searching for an oscillating massive scalar field as a dark matter candidate using atomic hyperfine frequency comparisons. *Phys. Rev. Lett.* **117**, 061301 (2016).

7. 7.

Leefer, N., Gerhardus, A., Budker, D., Flambaum, V. V. & Stadnik, Y. V. Search for the effect of massive bodies on atomic spectra and constraints on Yukawa-type interactions of scalar particles. *Phys. Rev. Lett.* **117**, 271601 (2016).

8. 8.

Savalle, E. et al. Searching for dark matter with an unequal delay interferometer. *Phys. Rev. Lett.* **126**, 051301 (2021).

9. 9.

Abbott, B. P. et al. GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Phys. Rev. X* **9**, 031040 (2019).

10. 10.

Abbott, B. P. et al. GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Pys. Rev. X* **11**, 021053 (2021).

11. 11.

Bertone, G. et al. Gravitational wave probes of dark matter: challenges and opportunities. *SciPost Phys. Core* **3**, 7 (2020).

12. 12.

Grote, H. On the possibility of vacuum QED measurements with gravitational wave detectors. *Phys. Rev. D* **91**, 022002 (2015).

13. 13.

Chou, A. S. et al. First measurements of high frequency cross-spectra from a pair of large Michelson interferometers. *Phys. Rev. Lett.* **117**, 111102 (2016).

14. 14.

Verlinde, E. P. & Zurek, K. M. Observational signatures of quantum gravity in interferometers. *Phys. Lett. B* **822**, 136663 (2021).

15. 15.

Vermeulen, S. M. et al. An experiment for observing quantum gravity phenomena using twin table-top 3D interferometers. *Class. Quantum Gravity* **38**, 085008 (2021).

16. 16.

Grote, H. & Stadnik, Y. V. Novel signatures of dark matter in laser-interferometric gravitational-wave detectors. *Phys. Rev. Res.* **1**, 033187 (2019).

17. 17.

Pierce, A., Riles, K. & Zhao, Y. Searching for dark photon dark matter with gravitational wave detectors. *Phys. Rev. Lett.* **121**, 061102 (2018).

18. 18.

Hall, E. D., Adhikari, R. X., Frolov, V. V., Müller, H. & Pospelov, M. Laser interferometers as dark matter detectors. *Phys. Rev. D* **98**, 083019 (2018).

19. 19.

Guo, H.-K., Riles, K., Yang, F.-W. & Zhao, Y. Searching for dark photon dark matter in LIGO O1 data. *Commun. Phys.* **2**, 155 (2019).

20. 20.

Read, J. I. The local dark matter density. *J. Phys. G Nucl. Part. Phys.* **41**, 063101 (2014).

21. 21.

Derevianko, A. Detecting dark matter waves with precision measurement tools. *Phys. Rev. A* **97**, 042506 (2018).

22. 22.

Flacke, T., Frugueule, C., Fuchs, E., Gupta, R. S. & Perez, G. Phenomenology of relaxion-Higgs mixing. *J. High Energy Phys.* **2017**, 50 (2017).

23. 23.

Banerjee, A., Kim, H. & Perez, G. Coherent relaxion dark matter. *Phys. Rev. D* **100**, 115026 (2019).

24. 24.

Banerjee, A., Budker, D., Eby, J., Kim, H. & Perez, G. Relaxion stars and their detection via atomic physics. *Commun. Phys.*, **3**, 1 (2020).

25. 25.

Ringwald, A. Exploring the role of axions and other WISPs in the dark universe. *Phys. Dark Universe* **1**, 116-135 (2012).

26. 26.

Hees, A., Minazzoli, O., Savalle, E., Stadnik, Y. V. & Wolf, P. Violation of the equivalence principle from light scalar dark matter. *Phys. Rev. D* **98**, 064051 (2018).

27. 27.

Damour, T. & Polyakov, A. M. The string dilation and a least coupling principle. *Nucl. Phys. B* **423**, 532–558 (1994).

28. 28.

Arvanitaki, A., Dimopoulos, S. & Van Tilburg, K. Sound of dark matter: searching for light scalars with resonant-mass detectors. *Phys. Rev. Lett.* **116**, 031102 (2016).

29. 29.

Damour, T. & Donoghue, J. F. Equivalence principle violations and couplings of a light dilaton. *Phys. Rev. D* **82**, 084033 (2010).

30. 30.

Geraci, A. A., Bradley, C., Gao, D., Weinstein, J. & Derevianko, A. Searching for ultra-light dark matter with optical cavities. *Phys. Rev. Lett.* **123**, 031304 (2019).

31. 31.

Lough, J. et al. First demonstration of 6 dB quantum noise reduction in a kilometer scale gravitational wave observatory. *Phys. Rev. Lett.* **126**, 041102 (2021).

32. 32.

Dooley, K. L. et al. GEO 600 and the GEO-HF upgrade program: successes and challenges. *Class. Quantum Gravity* **33**, 075009 (2016).

33. 33.

Tröbs, M. & Heinzel, G. Improved spectrum estimation from digitized time series on a logarithmic frequency axis. *Measurement* **39**, 120–129 (2006); corrigendum **42**, 170 (2009).

34. 34.

Miller, A. L. et al. Adapting a semi-coherent method to directly detect dark photon dark matter interacting with gravitational-wave interferometers. *Phys. Rev. D* **103**, 103002 (2021).

35. 35.

Abbott, B. P. et al. Properties of the binary neutron star merger GW170817. *Phys. Rev. X* **9**, 011001 (2019).

36. 36.

Freese, K., Lisanti, M. & Savage, C. Annual modulation of dark matter: A review. *Rev. Mod. Phys.* **85**, 1561–1581 (2013).

37. 37.

Wagner, T. A., Schlamminger, S., Gundlach, J. H. & Adelberger, E. G. Torsion-balance tests of the weak equivalence principle. *Class. Quantum Gravity* **29**, 184002 (2012).

38. 38.

Aharony, S. et al. Constraining rapidly oscillating scalar dark matter using dynamic decoupling. *Phys. Rev. D* **103**, 075017 (2019).

39. 39.  
Antypas, D. et al. Scalar dark matter in the radio-frequency band: atomic-spectroscopy search results. *Phys. Rev. Lett.* **123**, 141102 (2019).
40. 40.  
Branca, A. et al. Search for an ultralight scalar dark matter candidate with the AURIGA detector. *Phys. Rev. Lett.* **118**, 021302 (2017).
41. 41.  
Smith, G. L. et al. Short-range tests of the equivalence principle. *Phys. Rev. D* **61**, 022001 (1999).
42. 42.  
Schlamminger, S., Choi, K.-Y., Wagner, T. A., Gundlach, J. H. & Adelberger, E. G. Test of the equivalence principle using a rotating torsion balance. *Phys. Rev. Lett.* **100**, 041101 (2008).
43. 43.  
Bergé, J. et al. *MICROSCOPE* mission: first constraints on the violation of the weak equivalence principle by a light scalar dilaton. *Phys. Rev. Lett.* **120**, 141101 (2018).
44. 44.  
Pradyumna, S. T. et al. Twin beam quantum-enhanced correlated interferometry for testing fundamental physics. *Commun. Phys.* **3**, 104 (2020).
45. 45.  
Kennedy, C. J. et al. Precision metrology meets cosmology: improved constraints on ultralight dark matter from atom-cavity frequency comparisons. *Phys. Rev. Lett.* **125**, 201302 (2020).
46. 46.  
Savalle, E. et al. Novel approaches to dark-matter detection using space-time separated clocks. Preprint at <https://arxiv.org/abs/1902.07192> (2019).
47. 47.

Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73 (1967).

## Acknowledgements

We thank Y. Stadnik and Y. Michimura for discussion and comments on this work; D. Macleod and P. Hopkins for programming assistance; and M. Tröbs and G. Heinzel for permission to use their PSD code. We are grateful for support from the Science and Technology Facilities Council (STFC), grants ST/T006331/1, ST/I006285/1 and ST/L000946/1, the Leverhulme Trust, grant RPG-2019-022, and the universities of Cardiff and Glasgow in the UK, the Bundesministerium für Bildung und Forschung, the state of Lower Saxony in Germany, the Max Planck Society, Leibniz Universität Hannover and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC-2123 QuantumFrontiers 390837967. This work was also partly supported by DFG grant SFB/Transregio 7 on Gravitational Wave Astronomy. We further thank W. Grass for his years of expert infrastructure support for GEO600. This document has been assigned LIGO document number LIGO-P2100053.

## Author information

### Affiliations

1. Gravity Exploration Institute, Cardiff University, Cardiff, UK  
Sander M. Vermeulen, Philip Relton, Hartmut Grote & Vivien Raymond
2. Max Planck Institute for Gravitational Physics and Leibniz University Hannover, Hannover, Germany  
Christoph Affeldt, Fabio Bergamin, Aparna Bisht, Marc Brinkmann, Karsten Danzmann, Suresh Doravari, Volker Kringel, James Lough, Harald Lück, Moritz Mehmet, Nikhil Mukund, Séverin Nadji, Emil Schreiber, Kenneth A. Strain, Henning Vahlbruch, Michael Weinert, Benno Willke & Holger Wittel
3. School of Physics & Astronomy, University of Glasgow, Glasgow, UK  
Borja Sorazu & Kenneth A. Strain

### Contributions

S.M.V. and P.R. analysed the data and compiled the results; H.G. instigated this work and S.M.V. and H.G. wrote the manuscript; V.R. gave critical input to the analysis. C.A., J.L. and K.D. led the GEO600 instrument group during the period when data for this work were acquired. F.B., A.B., S.D., H.L., N.M., S.N., E.S., B.S., K.A.S., M.B., V.K., M.W. and H.W. worked on the instrument in different capacities required to achieve sensitivity and extended run duration; B.W. provided laser expertise and H.V. and M.M. built the squeezed-light source.

## Corresponding author

Correspondence to [Hartmut Grote](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 The spectral amplitude and SNR of a simulated dark matter signal and monochromatic sine wave as recovered from spectra created using different frequency bin widths \( \$\Delta f\_{\text{bin}} = 1/T\_{\text{DFT}}\$ \).](#)

The plotted recovered amplitude is normalized by the injected amplitude. The SNR ( $n_\sigma$ ) is measured as the difference between the signal amplitude and the noise amplitude divided by the standard deviation of the noise. The appearance of a maximum for the SNR as shown on the right is a direct consequence of both the decrease in the recovered amplitude of signals with limited coherence (as shown on the left) and the scaling of white Gaussian noise with increasing integration time. The plot on the left was produced by injecting a simulated dark matter (DM) signal and a perfect sine into a segment of GEO600 data and creating spectra using the modified

LPSD technique described above. The plot on the right was made by injecting the same signals into white Gaussian noise and creating spectra using Welch's method. Note that, for any single bin and for equal  $T_{\text{DFT}}$ , the spectral estimate obtained with the LPSD method (equation (8)) is mathematically equal to that obtained with Welch's method.

## Supplementary information

### [Peer Review File](#)

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### [Reprints and Permissions](#)

## About this article

### Cite this article

Vermeulen, S.M., Relton, P., Grote, H. *et al.* Direct limits for scalar field dark matter from a gravitational-wave detector. *Nature* **600**, 424–428 (2021).  
<https://doi.org/10.1038/s41586-021-04031-y>

- Received: 12 March 2021
- Accepted: 16 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021

- DOI: <https://doi.org/10.1038/s41586-021-04031-y>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04031-y>.

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 15 December 2021](#)

# Observation of Feshbach resonances between a single ion and ultracold atoms

- [Pascal Weckesser](#) [ORCID: orcid.org/0000-0003-3457-5841](#)<sup>1</sup>,
- [Fabian Thielemann](#)<sup>1</sup>,
- [Dariusz Wiater](#)<sup>2</sup>,
- [Agata Wojciechowska](#)<sup>2</sup>,
- [Leon Karpa](#)<sup>1,3</sup>,
- [Krzysztof Jachymski](#)<sup>2</sup>,
- [Michał Tomza](#) [ORCID: orcid.org/0000-0003-1792-8043](#)<sup>2</sup>,
- [Thomas Walker](#)<sup>1,4</sup> &
- [Tobias Schaetz](#)<sup>1,4</sup>

[Nature](#) volume **600**, pages 429–433 (2021)

- 1401 Accesses
- 83 Altmetric
- [Metrics details](#)

## Subjects

- [Atomic and molecular collision processes](#)
- [Quantum mechanics](#)
- [Ultracold gases](#)

# Abstract

The control of physical systems and their dynamics on the level of individual quanta underpins both fundamental science and quantum technologies. Trapped atomic and molecular systems, neutral<sup>1</sup> and charged<sup>2</sup>, are at the forefront of quantum science. Their extraordinary level of control is evidenced by numerous applications in quantum information processing<sup>3,4</sup> and quantum metrology<sup>5,6</sup>. Studies of the long-range interactions between these systems when combined in a hybrid atom–ion trap<sup>7,8</sup> have led to landmark results<sup>9,10,11,12,13,14,15,16,17,18,19</sup>. However, reaching the ultracold regime—where quantum mechanics dominates the interaction, for example, giving access to controllable scattering resonances<sup>20,21</sup>—has so far been elusive. Here we demonstrate Feshbach resonances between ions and atoms, using magnetically tunable interactions between <sup>138</sup>Ba<sup>+</sup> ions and <sup>6</sup>Li atoms. We tune the experimental parameters to probe different interaction processes—first, enhancing three-body reactions<sup>22,23</sup> and the related losses to identify the resonances and then making two-body interactions dominant to investigate the ion’s sympathetic cooling<sup>19</sup> in the ultracold atomic bath. Our results provide deeper insights into atom–ion interactions, giving access to complex many-body systems<sup>24,25,26,27</sup> and applications in experimental quantum simulation<sup>28,29,30</sup>.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.  
VAT will be added later in the checkout.  
Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

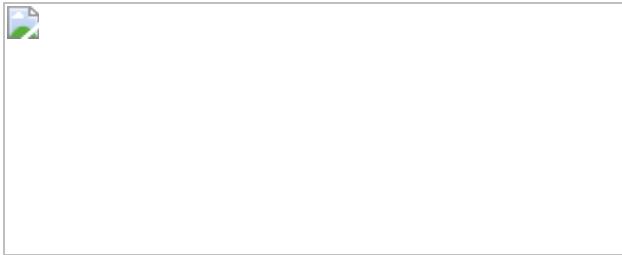
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

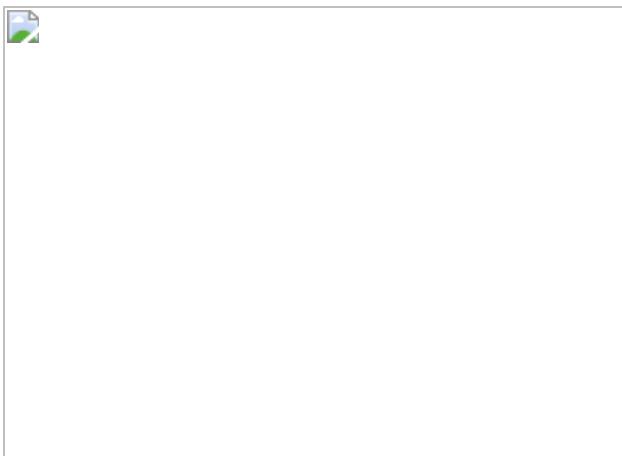
**Fig. 1: Experimental setup and concepts.**



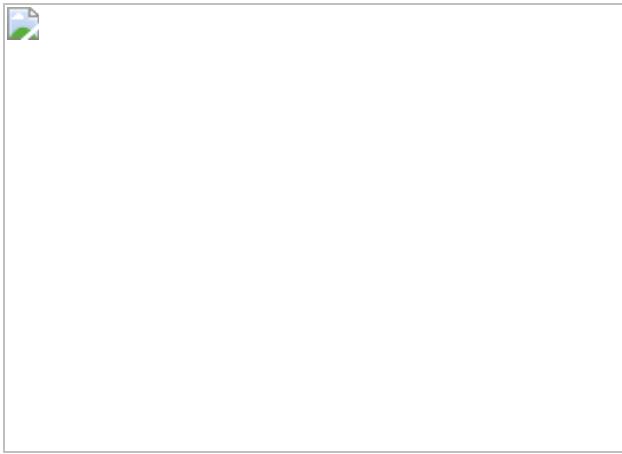
**Fig. 2: Detection of atom–ion Feshbach resonances by magnetic-field-dependent ion loss spectroscopy.**



**Fig. 3: Dependence of the ion loss rate on the atomic density at the Feshbach resonance at 296.31 G.**



**Fig. 4: Enhanced sympathetic cooling in the vicinity of the Feshbach resonance at 296.31 G.**



## Data availability

The experimental and theoretical data that support the findings of this study are available from the corresponding author upon request. [Source data](#) are provided with this paper.

## Code availability

The experimental data were analysed using JupyterLab and a self-written analysis script. Electronic structure calculations were performed with the MOLPRO package of ab initio programs<sup>67</sup> and multichannel quantum scattering calculations were realized with the extended version of QDYN program<sup>68</sup>. AMB model results were obtained and analysed with a self-written program and scripts in Mathematica and Python. The simulation results can be generated using the numerical methods described within Methods and the computer code developed, which are available upon request.

## References

1. 1.  
Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
2. 2.  
Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
3. 3.  
Wineland, D. Nobel lecture: Superposition, entanglement, and raising Schrödinger’s cat. *Rev. Mod. Phys.* **85**, 1103–1114 (2013).
4. 4.  
Saffman, M., Walker, T. & Mølmer, K. Quantum information with Rydberg atoms. *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
5. 5.

Katori, H. Optical lattice clocks and quantum metrology. *Nat. Photon.* **5**, 203–210 (2011).

6. 6.

Micke, P. et al. Coherent laser spectroscopy of highly charged ions using quantum logic. *Nature* **578**, 60–65 (2020).

7. 7.

Härter, A. & Hecker Denschlag, J. Cold atom–ion experiments in hybrid traps. *Contemp. Phys.* **55**, 33–45 (2014).

8. 8.

Tomza, M. et al. Cold hybrid ion–atom systems. *Rev. Mod. Phys.* **91**, 035001 (2019).

9. 9.

Grier, A., Cetina, M., Oručević, F. & Vuletić, V. Observation of cold collisions between trapped ions and trapped atoms. *Phys. Rev. Lett.* **102**, 223201 (2009).

10. 10.

Rellergert, W. G. et al. Measurement of a large chemical reaction rate between ultracold closed-shell  $^{40}\text{Ca}$  atoms and open-shell  $^{174}\text{Yb}^+$  ions held in a hybrid atom–ion trap. *Phys. Rev. Lett.* **107**, 243201 (2011).

11. 11.

Ratschbacher, L., Zipkes, C., Sias, C. & Köhl, M. Controlling chemical reactions of a single particle. *Nat. Phys.* **8**, 649–652 (2012).

12. 12.

Hall, F., Aymar, M., Raoult, M., Dulieu, O. & Willitsch, S. Light-assisted cold chemical reactions of barium ions with rubidium atoms.

*Mol. Phys.* **111**, 1683–1690 (2013).

13. 13.

Ratschbacher, L. et al. Decoherence of a single-ion qubit immersed in a spin-polarized atomic bath. *Phys. Rev. Lett.* **110**, 160402 (2013).

14. 14.

Meir, Z. et al. Dynamics of a ground-state cooled ion colliding with ultracold atoms. *Phys. Rev. Lett.* **117**, 243401 (2016).

15. 15.

Saito, R. et al. Characterization of charge-exchange collisions between ultracold  ${}^6\text{Li}$  atoms and  ${}^{40}\text{Ca}^+$  ions. *Phys. Rev. A* **95**, 032709 (2017).

16. 16.

Joger, J. et al. Observation of collisions between cold Li atoms and  $\text{Yb}^+$  ions. *Phys. Rev. A* **96**, 030703 (2017).

17. 17.

Fürst, H. et al. Dynamics of a single ion-spin impurity in a spin-polarized atomic bath. *Phys. Rev. A* **98**, 012713 (2018).

18. 18.

Sikorsky, T., Meir, Z., Ben-Shlomi, R., Akerman, N. & Ozeri, R. Spin-controlled atom–ion chemistry. *Nat. Commun.* **9**, 920 (2018).

19. 19.

Feldker, T. et al. Buffer gas cooling of a trapped ion to the quantum regime. *Nat. Phys.* **16**, 413–416 (2020).

20. 20.

Idziaszek, Z., Simoni, A., Calarco, T. & Julienne, P. Multichannel quantum-defect theory for ultracold atom–ion collisions. *New J. Phys.* **13**, 083005 (2011).

21. 21.

Tomza, M., Koch, C. & Moszynski, R. Cold interactions between an  $\text{Yb}^+$  ion and a Li atom: prospects for sympathetic cooling, radiative association, and Feshbach resonances. *Phys. Rev. A* **91**, 042706 (2015).

22. 22.

Härter, A. et al. Single ion as a three-body reaction center in an ultracold atomic gas. *Phys. Rev. Lett.* **109**, 123201 (2012).

23. 23.

Krükow, A., Mohammadi, A., Härter, A. & Denschlag, J. H. Reactive two-body and three-body collisions of  $\text{Ba}^+$  in an ultracold Rb gas. *Phys. Rev. A* **94**, 030701 (2016).

24. 24.

Cote, R., Kharchenko, V. & Lukin, M. Mesoscopic molecular ions in Bose–Einstein condensates. *Phys. Rev. Lett.* **89**, 093001 (2002).

25. 25.

Casteels, W., Tempere, J. & Devreese, J. Polaronic properties of an ion in a Bose–Einstein condensate in the strong-coupling limit. *J. Low Temp. Phys.* **162**, 266–273 (2011).

26. 26.

Jachymski, K. & Negretti, A. Quantum simulation of extended polaron models using compound atom–ion systems. *Phys. Rev. Res.* **2**, 033326 (2020).

27. 27.

Hirzler, H. et al. Controlling the nature of a charged impurity in a bath of Feshbach dimers. *Phys. Rev. Res.* **2**, 033232 (2020).

28. 28.

Doerk, H., Idziaszek, Z. & Calarco, T. Atom–ion quantum gate. *Phys. Rev. A* **81**, 012708 (2010).

29. 29.

Gerritsma, R. et al. Bosonic Josephson junction controlled by a single trapped ion. *Phys. Rev. Lett.* **109**, 080402 (2012).

30. 30.

Bissbort, U. et al. Emulating solid-state physics with a hybrid system of ultracold ions and atoms. *Phys. Rev. Lett.* **111**, 080501 (2013).

31. 31.

Inouye, S. et al. Observation of Feshbach resonances in a Bose–Einstein condensate. *Nature* **392**, 151–154 (1998).

32. 32.

Inouye, S. et al. Observation of heteronuclear Feshbach resonances in a mixture of bosons and fermions. *Phys. Rev. Lett.* **93**, 183201 (2004).

33. 33.

Barbé, V. et al. Observation of Feshbach resonances between alkali and closed-shell atoms. *Nat. Phys.* **14**, 881–884 (2018).

34. 34.

Yang, H. et al. Observation of magnetically tunable Feshbach resonances in ultracold  $^{23}\text{Na}^{40}\text{K} + ^{40}\text{K}$  collisions. *Science* **363**, 261–264 (2019).

35. 35.

Aikawa, K. et al. Bose–Einstein condensation of erbium. *Phys. Rev. Lett.* **108**, 210401 (2012).

36. 36.

Durastante, G. et al. Feshbach resonances in an erbium–dysprosium dipolar mixture. *Phys. Rev. A* **102**, 033330 (2020).

37. 37.

Cetina, M., Grier, A. & Vuletić, V. Micromotion-induced limit to atom–ion sympathetic cooling in Paul traps. *Phys. Rev. Lett.* **109**, 253201 (2012).

38. 38.

Chen, K., Sullivan, S. T. & Hudson, E. R. Neutral gas sympathetic cooling of an ion in a Paul trap. *Phys. Rev. Lett.* **112**, 143009 (2014).

39. 39.

Höltkemeier, B., Weckesser, P., López-Carrera, H. & Weidemüller, M. Buffer-gas cooling of a single ion in a multipole radio frequency trap beyond the critical mass ratio. *Phys. Rev. Lett.* **116**, 233003 (2016).

40. 40.

Rouse, I. & Willitsch, S. Superstatistical energy distributions of an ion in an ultracold buffer gas. *Phys. Rev. Lett.* **118**, 143401 (2017).

41. 41.

Lambrecht, A. et al. Long lifetimes and effective isolation of ions in optical and electrostatic traps. *Nat. Photon.* **11**, 704–707 (2017).

42. 42.

Schmidt, J. et al. Optical trapping of ion Coulomb crystals. *Phys. Rev. X* **8**, 021028 (2018).

43. 43.

Weckesser, P. et al. Trapping, shaping, and isolating of an ion Coulomb crystal via state-selective optical potentials. *Phys. Rev. A* **103**, 013112 (2021).

44. 44.

Schmidt, J., Weckesser, P., Thielemann, F., Schaetz, T. & Karpa, L. Optical traps for sympathetic cooling of ions with ultracold neutral atoms. *Phys. Rev. Lett.* **124**, 053402 (2020).

45. 45.

Schneider, C., Enderlein, M., Huber, T., Dürr, S. & Schaetz, T. Influence of static electric fields on an optical ion trap. *Phys. Rev. A* **85**, 013422 (2012).

46. 46.

Ticknor, C., Regal, C., Jin, D. & Bohn, J. Multiplet structure of Feshbach resonances in nonzero partial waves. *Phys. Rev. A* **69**, 042712 (2004).

47. 47.

Cui, Y. et al. Observation of broad *d*-wave Feshbach resonances with a triplet structure. *Phys. Rev. Lett.* **119**, 203402 (2017).

48. 48.

Maier, T. et al. Emergence of chaotic scattering in ultracold Er and Dy. *Phys. Rev. X* **5**, 041029 (2015).

49. 49.

Köhler, T., Góral, K. & Julienne, P. Production of cold molecules via magnetically tunable Feshbach resonances. *Rev. Mod. Phys.* **78**, 1311 (2006).

50. 50.

Wester, R. Radiofrequency multipole traps: tools for spectroscopy and dynamics of cold molecular ions. *J. Phys. B* **42**, 154001 (2009).

51. 51.

Jochim, S. et al. Magnetic field control of elastic scattering in a cold gas of fermionic lithium atoms. *Phys. Rev. Lett.* **89**, 273202 (2002).

52. 52.

Luo, L. et al. Evaporative cooling of unitary Fermi gas mixtures in optical traps. *New J. Phys.* **8**, 213 (2006).

53. 53.

Ketterle, W. & Martin W. Z. Making, probing and understanding ultracold Fermi gases. *La Rivista del Nuovo Cimento* **31**, 247–422 (2008).

54. 54.

Grimm, R., Weidemüller, M. & Ovchinnikov, Y. in *Advances in Atomic, Molecular, and Optical Physics* (eds Bederson, B. & Walther, H.) Vol. 42, 95–170 (Elsevier, 2000).

55. 55.

Karpa, L. *Trapping Single Ions and Coulomb Crystals with Light Fields* (Springer, 2019).

56. 56.

Maxwell, J. *A Treatise on Electricity and Magnetism* Vol. 1  
(Clarendon, 1873).

57. 57.

Krükow, A. et al. Energy scaling of cold atom–atom–ion three-body recombination. *Phys. Rev. Lett.* **116**, 193201 (2016).

58. 58.

Pérez-Ros, J. & Greene, C. H. Universal temperature dependence of the ion–neutral–neutral three-body recombination rate. *Phys. Rev. A* **98**, 062707 (2018).

59. 59.

Pérez-Ros, J. & Greene, C. Communication: Classical threshold law for ion–neutral–neutral three-body recombination. *J. Chem. Phys.* **143**, 041105 (2015).

60. 60.

Mohammadi, A. et al. Life and death of a cold BaRb<sup>+</sup> molecule inside an ultracold cloud of Rb atoms. *Phys. Rev. Res.* **3**, 013196 (2021).

61. 61.

Breit, G. & Rabi, I. Measurement of nuclear spin. *Phys. Rev.* **38**, 2082–2083 (1931).

62. 62.

Zhang, J. et al. *P*-wave Feshbach resonances of ultracold <sup>6</sup>Li. *Phys. Rev. A* **70**, 030702 (2004).

63. 63.

Mies, F., Williams, C., Julienne, P. & Krauss, M. Estimating bounds on collisional relaxation rates of spin-polarized <sup>87</sup>Rb atoms at ultracold

temperatures. *J. Res. Natl Inst. Stand. Technol.* **101**, 521–535 (1996).

64. 64.

Tscherbul, T., Brumer, P. & Buchachenko, A. Spin–orbit interactions and quantum spin dynamics in cold ion–atom collisions. *Phys. Rev. Lett.* **117**, 143201 (2016).

65. 65.

Wille, E. et al. Exploring an ultracold Fermi–Fermi mixture: interspecies Feshbach resonances and scattering properties of  ${}^6\text{Li}$  and  ${}^{40}\text{K}$ . *Phys. Rev. Lett.* **100**, 053201 (2008).

66. 66.

Tiecke, T., Goosen, M., Walraven, J. & Kokkelmans, S. Asymptotic-bound-state model for Feshbach resonances. *Phys. Rev. A* **82**, 042712 (2010).

67. 67.

Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R. & Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2**, 242–253 (2012).

68. 68.

Goerz, M. H., Reich, D. M., Tomza, M. & Koch, C. QDYN, version 1.0, a program package for quantum dynamics and control (2015); <https://qdyn-library.net>

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant number 648330) and was supported by the Georg H. Endress foundation. P.W., F.T. and T.S. acknowledge support from the DFG

within the GRK 2079/1 programme. P.W. gratefully acknowledges financial support from the Studienstiftung des deutschen Volkes. L.K. is grateful for financial support from Marie Curie Actions. D.W., A.W. and M.T. acknowledge the financial support from the National Science Centre Poland (grant numbers 2016/23/B/ST4/03231 and 2020/36/T/ST2/00591) and Foundation for Polish Science within the First Team programme co-financed by the European Union under the European Regional Development Fund. K.J. acknowledges support from the Polish National Agency for Academic Exchange (NAWA) via the Polish Returns 2019 programme. The computational part was partially supported by the PL-Grid Infrastructure. We thank O. Dulieu for discussions. We thank M. Debatin for building the experimental apparatus.

## Author information

### Affiliations

1. Albert-Ludwigs-Universität Freiburg, Physikalisches Institut, Freiburg, Germany

Pascal Weckesser, Fabian Thielemann, Leon Karpa, Thomas Walker & Tobias Schaetz

2. Faculty of Physics, University of Warsaw, Warsaw, Poland

Dariusz Wiater, Agata Wojciechowska, Krzysztof Jachymski & Michał Tomza

3. Leibniz University Hannover, Institute of Quantum Optics, Hannover, Germany

Leon Karpa

4. EUCOR Centre for Quantum Science and Quantum Computing, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

Thomas Walker & Tobias Schaetz

## Contributions

T.S. conceived the experiments. P.W. and F.T. contributed equally to the construction of the setup, carrying out of the experiments, discussion of the results and analysis of the data and were supported by L.K. and T.W. D.W., A.W., K.J. and M.T. performed theoretical calculations and analysis supervised by M.T. T.S. supervised the work. P.W. and T.S. wrote the manuscript with contributions from T.W., K.J. and M.T. All authors worked on the interpretation of the data and contributed to the final manuscript.

## Corresponding author

Correspondence to [Pascal Weckesser](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Takashi Mukaiyama and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Electronic level scheme of \$^{138}\text{Ba}^+\(I=0\)\$ .](#)

We label the relevant electronic dipole transitions with their respective wavelength  $\lambda$  and natural linewidth  $\Gamma$ . We Doppler cool the ion

driving the  $6\{\rm{S}\}_{1/2} \leftrightarrow 6\{\rm{P}\}_{1/2}$ ) and  $(5\{\rm{D}\}_{3/2} \leftrightarrow 6\{\rm{P}\}_{1/2})$  transition.

Inelastic losses, such as TBR followed by light-assisted dissociation, can partially result in the ion's population of the  $5D_{5/2}$  manifold<sup>60</sup>. We detect these events through optical pumping with 614 nm laser light, followed by fluorescence detection while Doppler cooling.

## Extended Data Fig. 2 Time-dependent $\text{Ba}^+$ and Li loss for variable atomic density $n$ around 296.31G.

(upper) Ion survival probability while interacting with the  ${}^6\text{Li}$  cloud for various atomic densities  $n$ . Data points are an average of at least 20 independent experimental realizations. Error bars denote the upper bound of the  $1\sigma$ -confidence interval of the underlying binomial distribution. The solid lines are exponential fits  $\langle(\langle e \rangle^{-\langle t \rangle_{\{\rm{int}\}}}) \backslash \text{varGamma}_{\{\rm{Loss}\}} \rangle$  to the respective data. The fit results and the respective error bars are illustrated as density-dependent loss rate  $\langle(\langle \text{varGamma} \rangle_{\{\rm{Loss}\}})(n) \rangle$  in Fig. 3. (lower) Normalized number of remaining  ${}^6\text{Li}$  atoms interacting with a single  ${}^{138}\text{Ba}^+$  ion in dependence on  $\langle\langle t \rangle_{\{\rm{int}\}} \rangle$ . The markers and the respective colors indicate the association to the data in the upper graph. Li atoms are removed from the XODT by either spin-changing collisions or elastic atom-ion interactions. For the presented analysis, we exclude experiments resulting in ion loss, to avoid systematic errors by inelastic collisions. To mitigate the density uncertainty due to the decay of the atom number, we choose interaction durations resulting in maximal atom loss of  $\langle(\lessapprox 10\%) \rangle$ . We further indicate the atom number evolution in absence of interaction (black circles and solid line).

[Source data](#)

## Extended Data Fig. 3 Potential energy curves for a $\text{Ba}^+$ ion interacting with a Li atom.

The interaction between ground-state  $\text{Ba}^+$  ion and Li atom results in two molecular electronic states of the singlet  $(\{\text{X}\}^1 \{\Sigma^+\})$  (solid black line) and triplet  $(\{\text{a}\}^3 \{\Sigma^+\})$  (solid red line) symmetries. The excited molecular electronic state of the triplet  $(\{\text{b}\}^3 \{\text{Pi}\})$  symmetry (dashed red line) originates from the interaction of  $\text{Ba}^+$  ion in the lowest excited  $(\{\}^2 \text{D})$  state and ground-state Li atom and crosses the  $(\{\text{a}\}^3 \{\Sigma^+\})$  state at a small interatomic distance. This crossing combined with SOC between  $(\{\text{a}\}^3 \{\Sigma^+\})$  and  $(\{\text{b}\}^3 \{\text{Pi}\})$  states results in large second-order SOC in the collision channels responsible for the observed Feshbach resonances. Further illustrated is a possible photodissociation transition induced by our xODT laser operated at 1064 nm. The observed TBR might result in the formation of weakly-bound molecular ions. These can couple by laser light to higher energetic asymptotes, resulting in the population of excited states, including  $\text{Ba}(\{\}^+ (5\{\text{rm}\{\text{D}\}\})_{-5/2}))$ .

## Extended Data Table 1 List of observed Feshbach resonances for the entrance channel

## Source data

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 2](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article

## Cite this article

Weckesser, P., Thielemann, F., Wiater, D. *et al.* Observation of Feshbach resonances between a single ion and ultracold atoms. *Nature* **600**, 429–433 (2021). <https://doi.org/10.1038/s41586-021-04112-y>

- Received: 19 May 2021
- Accepted: 08 October 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04112-y>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04112-y>.

- Article
- [Published: 15 December 2021](#)

# Detecting spins by their fluorescence with a microwave photon counter

- [Emanuele Albertinale](#)<sup>1</sup>,
- [Léo Balembois](#)<sup>1</sup>,
- [Eric Billaud](#)<sup>1</sup>,
- [Vishal Ranjan](#)<sup>1</sup>,
- [Daniel Flanigan](#)<sup>1</sup>,
- [Thomas Schenkel](#)<sup>2</sup>,
- [Daniel Estève](#) [ORCID: orcid.org/0000-0003-4089-4582](#)<sup>1</sup>,
- [Denis Vion](#)<sup>1</sup>,
- [Patrice Bertet](#) [ORCID: orcid.org/0000-0002-5427-7254](#)<sup>1</sup> &
- [Emmanuel Flurin](#) [ORCID: orcid.org/0000-0002-1367-8967](#)<sup>1</sup>

[Nature](#) volume **600**, pages 434–438 (2021)

- 1362 Accesses
- 14 Altmetric
- [Metrics details](#)

## Subjects

- [Electronics, photonics and device physics](#)
- [Quantum physics](#)

# Abstract

Quantum emitters respond to resonant illumination by radiating part of the absorbed energy. A component of this radiation field is phase coherent with the driving tone, whereas another component is incoherent and consists of spontaneously emitted photons, forming the fluorescence signal<sup>1</sup>. Atoms, molecules and colour centres are routinely detected by their fluorescence at optical frequencies, with important applications in quantum technology<sup>2,3</sup> and microscopy<sup>4,5,6,7</sup>. By contrast, electron spins are usually detected by the phase-coherent echoes that they emit in response to microwave driving pulses<sup>8</sup>. The incoherent part of their radiation—a stream of microwave photons spontaneously emitted upon individual spin relaxation events—has not been observed so far because of the low spin radiative decay rate and of the lack of single microwave photon detectors (SMPDs). Here using superconducting quantum devices, we demonstrate the detection of a small ensemble of donor spins in silicon by their fluorescence at microwave frequencies and millikelvin temperatures. We enhance their radiative decay rate by coupling them to a high-quality-factor and small-mode-volume superconducting resonator<sup>9</sup>, and we connect the device output to a newly developed SMPD<sup>10</sup> based on a superconducting qubit. In addition, we show that the SMPD can be used to detect spin echoes and that standard spin characterization measurements (Rabi nutation and spectroscopy) can be achieved with both echo and fluorescence detection. We discuss the potential of SMPD detection as a method for magnetic resonance spectroscopy of small numbers of spins.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

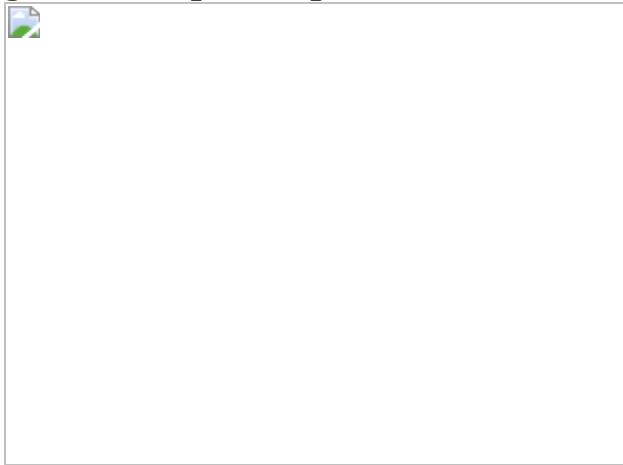
[Rent or Buy](#)

All prices are NET prices.

**Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

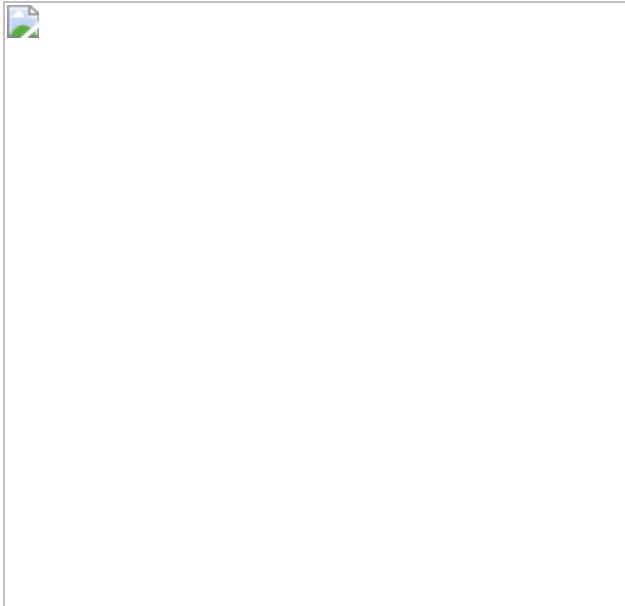
**Fig. 1: Principle of spin detection with a photon counter.**



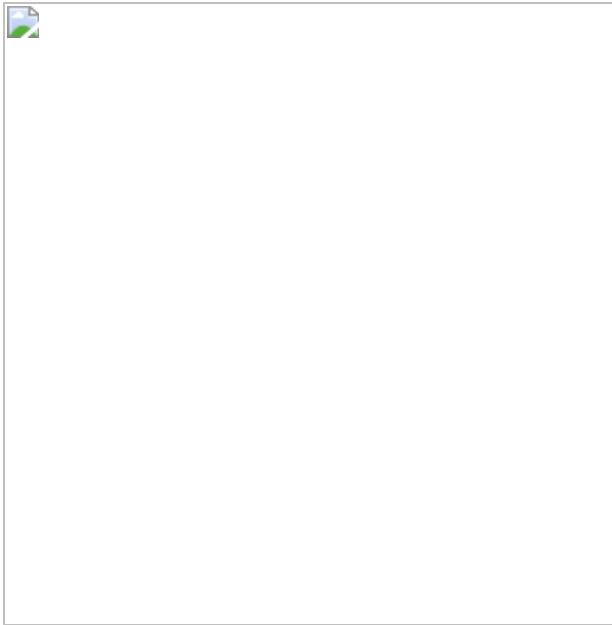
**Fig. 2: Detection of spin relaxation by photon counting.**



**Fig. 3: Detection of spin echo by photon counting.**



**Fig. 4: Spin ensemble characterization by photon counting.**



## Data availability

The datasets acquired and analysed during the current study are available from the corresponding author on reasonable request.

## References

1. 1.

Mandel, L. & Wolf, E. *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, 1995).

2. 2.

Kimble, H. J., Dagenais, M. & Mandel, L. Photon antibunching in resonance fluorescence. *Phys. Rev. Lett.* **39**, 691–695 (1977).

3. 3.

Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: a vision for the road ahead. *Science* **362**, eaam9288 (2018).

4. 4.

Orrit, M. & Bernard, J. Single pentacene molecules detected by fluorescence excitation in a *p*-terphenyl crystal. *Phys. Rev. Lett.* **65**, 2716–2719 (1990).

5. 5.

Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl Acad. Sci. USA* **97**, 8206–8210 (2000).

6. 6.

Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

7. 7.

Bruschini, C., Homulle, H., Antolovic, I. M., Burri, S. & Charbon, E. Single-photon avalanche diode imagers in biophotonics: review and outlook. *Light Sci. Appl.* **8**, 87 (2019).

8. 8.

Schweiger, A. & Jeschke, G. *Principles of Pulse Electron Paramagnetic Resonance* (Oxford Univ. Press, 2001).

9. 9.

Bienfait, A. et al. Controlling spin relaxation with a cavity. *Nature* **531**, 74–77 (2016).

10. 10.

Lescanne, R. et al. Irreversible qubit-photon coupling for the detection of itinerant microwave photons. *Phys. Rev.* **10**, 021038 (2020).

11. 11.

Roy, A. & Devoret, M. Introduction to parametric amplification of quantum signals with Josephson circuits. *C. R. Phys.* **17**, 740–755 (2016).

12. 12.

Bienfait, A. et al. Reaching the quantum limit of sensitivity in electron spin resonance. *Nat. Nanotechnol.* **11**, 253–257 (2016).

13. 13.

Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 62320 (2004).

14. 14.

Lamoreaux, S. K., van Bibber, K. A., Lehnert, K. W. & Carosi, G. Analysis of single-photon and linear amplifier detectors for microwave cavity dark matter axion searches. *Phys. Rev. D* **88**, 035020 (2013).

15. 15.

Gleyzes, S. et al. Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature* **446**, 297–300 (2007).

16. 16.

Chen, Y.-F. et al. Microwave photon counter based on Josephson junctions. *Phys. Rev. Lett.* **107**, 217401 (2011).

17. 17.

Inomata, K. et al. Single microwave-photon detector using an artificial lambda-type three-level system. *Nat. Commun.* **7**, 12303 (2016).

18. 18.

Narla, A. et al. Robust concurrent remote entanglement between two superconducting qubits. *Phys. Rev.* **6**, 031036 (2016).

19. 19.

Besse, J.-C. et al. Single-shot quantum nondemolition detection of individual itinerant microwave photons. *Phys. Rev.* **8**, 021003 (2018).

20. 20.

Kono, S., Koshino, K., Tabuchi, Y., Noguchi, A. & Nakamura, Y. Quantum non-demolition detection of an itinerant microwave photon. *Nat. Phys.* **14**, 546–549 (2018).

21. 21.

Astner, T. et al. Solid-state electron spin lifetime limited by phononic vacuum modes. *Nat. Mater.* **17**, 313–317 (2018).

22. 22.

Eichler, C., Sigillito, A. J., Lyon, S. A. & Petta, J. R. Electron spin resonance at the level of  $10^4$  spins using low impedance superconducting resonators. *Phys. Rev. Lett.* **118**, 037701 (2017).

23. 23.

Sleator, T., Hahn, E. L., Hilbert, C. & Clarke, J. Nuclear-spin noise. *Phys. Rev. Lett.* **55**, 1742 (1985).

24. 24.

McCoy, M. A. & Ernst, R. R. Nuclear spin noise at room temperature. *Chem. Phys. Lett.* **159**, 587–593 (1989).

25. 25.

Tyryshkin, A. M. et al. Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).

26. 26.

Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).

27. 27.

Palacios-Laloy, A. et al. Tunable resonators for quantum circuits. *J. Low Temp. Phys.* **151**, 1034–1042 (2008).

28. 28.

Gao, J. et al. Experimental evidence for a surface distribution of two-level systems in superconducting lithographed microwave resonators. *Appl. Phys. Lett.* **92**, 152505 (2008).

29. 29.

Zhu, X. et al. Coherent coupling of a superconducting flux qubit to an electron spin ensemble in diamond. *Nature* **478**, 221–224 (2011).

30. 30.

Kubo, Y. et al. Electron spin resonance detected by a super conducting qubit. *Phys. Rev. B* **86**, 064514 (2012).

31. 31.

Toida, H. et al. Electron paramagnetic resonance spectroscopy using a single artificial atom. *Commun. Phys.* **2**, 33 (2019).

32. 32.

Budoyo, R., Kakuyanagi, K., Toida, H., Matsuzaki, Y. & Saito, S. Electron spin resonance with up to 20 spin sensitivity measured using a superconducting flux qubit. *Appl. Phys. Lett.* **116**, 194001 (2020).

33. 33.

Ranjan, V. et al. Electron spin resonance spectroscopy with femtoliter detection volume. *Appl. Phys. Lett.* **116**, 184002 (2020).

34. 34.

Abella, I. D., Kurnit, N. A. & Hartmann, S. R. Photon echoes. *Phys. Rev.* **141**, 391–406 (1966).

35. 35.

Ranjan, V. et al. Pulsed electron spin resonance spectroscopy in the Purcell regime. *J. Magn. Reson.* **310**, 106662 (2020).

36. 36.

Royer, B., Grimsmo, A. L., Choquette-Poitevin, A. & Blais, A. Itinerant microwave photon detector. *Phys. Rev. Lett.* **120**, 203602 (2018).

37. 37.

Grimsmo, A. L. et al. Quantum metamaterial for broadband detection of single microwave photons. *Phys. Rev. Appl.* **15**, 034074 (2021).

38. 38.

Kokkoniemi, R. et al. Bolometer operating at the threshold for circuit quantum electrodynamics. *Nature* **586**, 47–51 (2020).

39. 39.

Zhou, X. et al. High-gain weakly nonlinear flux-modulated Josephson parametric amplifier using a SQUID array. *Phys. Rev. B* **89**, 214517 (2014).

40. 40.

Dunsworth, A. et al. Characterization and reduction of capacitive loss induced by sub-micron Josephson junction fabrication in

superconducting qubits. *Appl. Phys. Lett.* **111**, 022601 (2017).

41. 41.

Bruno, A. et al. Reducing intrinsic loss in superconducting resonators by surface treatment and deep etching of silicon substrates. *Appl. Phys. Lett.* **106**, 182601 (2015).

42. 42.

Calusine, G. et al. Analysis and mitigation of interface losses in trenched superconducting coplanar waveguide resonators. *Appl. Phys. Lett.* **112**, 062601 (2018).

43. 43.

Pla, J. et al. Strain-induced spin-resonance shifts in silicon devices. *Phys. Rev. Appl.* **9**, 044014 (2018).

44. 44.

Gambetta, J. et al. Qubit-photon interactions in a cavity: measurement-induced dephasing and number splitting. *Phys. Rev. A* **74**, 042318 (2006).

45. 45.

Probst, S. et al. Inductive-detection electron-spin resonance spectroscopy with 65 spins/ $\sqrt{\text{Hz}}$  sensitivity. *Appl. Phys. Lett.* **111**, 202604 (2017).

## Acknowledgements

We acknowledge technical support from P. Sénat, D. Duet, P.-F. Orfila and S. Delprat, and are grateful for discussions within the Quantronics group. This project has received funding from the European Union Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement number 765267 (QuSCO). E.F. has received funding from the

Agence Nationale de la Recherche (ANR) under grant agreement number DARKWADOR:ANR-19-CE47-0004. We acknowledge support from the Agence Nationale de la Recherche (ANR) through the Chaire Industrielle NASNIQ under contract ANR-17-CHIN-0001 cofunded by Atos, and of the Région Ile-de-France through the DIM SIRTEQ (REIMIC project).

## Author information

### Affiliations

1. Université Paris-Saclay, CEA, CNRS, SPEC, Gif-sur-Yvette, France

Emanuele Albertinale, Léo Balembois, Eric Billaud, Vishal Ranjan, Daniel Flanigan, Daniel Estève, Denis Vion, Patrice Bertet & Emmanuel Flurin

2. Accelerator Technology and Applied Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Thomas Schenkel

### Contributions

E.A., P.B. and E.F. designed the experiment. T.S. provided the bismuth-implanted isotopically purified silicon sample, on which V.R. fabricated the aluminium resonator. E.A. designed and fabricated the SMPD with the help of D.V. and E.F. E.A., V.R. and E.F. performed the measurements, with help from L.B., D.F. and P.B. E.A., P.B. and E.F. analysed the data. E.A., E.B. and V.R. performed the simulations. E.A., P.B. and E.F. wrote the manuscript. D.F., D.V. and D.E. contributed useful input to the manuscript.

### Corresponding author

Correspondence to [Emmanuel Flurin](#).

### Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Gunnar Jeschke and Stefan Putz for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Schematic of the setup.

Wiring diagram at room temperature and cryogenic temperature extensively described in the methods.

### Extended Data Fig. 2 ESR spectroscopy of the spin ensemble.

**a**, Measured (blue dots) and simulated (solid line) integrated echo as a function of the delay  $\tau$  between the inversion  $\pi$  pulse and the Hahn echo sequence. An exponential fit (not shown) yield a characteristic decay time  $\langle\{T\}_1\rangle=300\pm 10\,\text{ms}$ . **b**, Measured (blue dots) and simulated (solid line) integrated echo as a function of the delay  $\tau$  between  $\pi/2$  and  $\pi$  pulses of the Hahn echo sequence. An exponential fit (not showed) yields a characteristic decay time  $T_2 = 2.7 \text{ ms}$ . **c**, Measured (blue dots) and simulated (solid line) integrated echo as a function of the amplitude  $A$  of the  $\pi$  pulse of the Hahn echo sequence revealing Rabi oscillations. **d**, Measured integrated echo (blue dots) as a function of the in-plane magnetic field  $B_0$  used to tune the spin ensemble frequency.

### Extended Data Fig. 3 SMPD characterization.

**a**, Measured (blue dots) and fitted (orange solid line) complex reflection coefficient  $r$  of the buffer resonator at working point  $\omega_b/2\pi = 6.946 \omega_{\text{G}} \omega_{\text{H}} \omega_z$ . The fitting function takes into account flux noise of the SQUID enabling the tuning of the resonator. **b, d**, Probability  $P_e$  of finding the qubit in its excited state (color scale) as function of the amplitude  $A_{\text{pump}}$  and frequency  $\omega_{\text{pump}}/2\pi$  of the pump activating the parametric process of photo-detection. When no photon is impinging, the buffer resonator is close to its vacuum state (**b**) no parametric process is activated and the qubit is mostly in its ground state ( $P_e \approx 0$ ); in contrast, when photons are injected (**d**) the parametric process is activated at pump frequencies for which the conservation of energy is respected. The quadratic dependence of the pump activation frequency on the pump amplitude is due to the Stark shift of the qubit frequency for increasing pump power. **c**, Measured (blue dots) efficiency of detection  $\eta$  at  $\omega_b/2\pi = 6.946 \omega_{\text{G}} \omega_{\text{H}} \omega_z$ , as a function of the input photon frequency  $\omega_{\text{photon}}/2\pi$ . From the fit (orange solid line), obtained with a model of two coupled cavities, we extract a bandwidth  $\Delta_{\text{det}}/2\pi \approx 2.1 \omega_M \omega_H \omega_z$ .

## Extended Data Fig. 4 SMPD performance.

**a, b**, Probability  $p(V)$  of measuring the average quadrature voltage  $V$  when probing the buffer resonator for qubit readout, when a pulse is applied (**b**) or not (**a**) to the qubit prior to the measurement. Dashed line indicates the readout threshold, chosen to minimise the ratio  $\nu_{dc}/\eta_d$ , each measure falling on the left (resp. right) is associated to the qubit being in its ground (resp. excited) state. **c**, Measured (blue dots) and fitted (orange solid line) probability  $P_e$  of finding the qubit in its excited state as function of time  $T$  after the reset sequence, showing out-of-equilibrium qubit excited population reaching thermal equilibrium on a timescale  $\approx T_1$ . Black dashed line at  $t_D = 5 \mu_s$  marks the point at which detector is operated. **d**, Measured (blue dots) and fitted (orange solid line) detector efficiency  $\eta$

$\}_{\{ \{ \rm{d} \} \}} \})$ ) as a function of the duration of the detection step  $\langle \{ t \}_{\{ \{ \rm{D} \} \}} \rangle$ ). Fit model takes into account bandwidth-limited detection efficiency for short detection windows and  $T_1$ -decay effect for increasing  $\langle \{ t \}_{\{ \{ \rm{d} \} \}} \rangle$ ). Black dashed line at  $\langle \{ t \}_{\{ \{ \rm{D} \} \}} \rangle = 5$ ,  $\langle \{ \rm{\mu} \} \}_{\{ \{ \rm{s} \} \}}$ ) marks the point at which detector is operated, to optimise the photo-detected echo signal. **e**, Dark count rate  $\langle \{ \nu \}_{\{ \{ \rm{dc} \} \}} \rangle$  as a function of time when a train of photo-detection sequences is continuously sent to the fridge. An equilibrium value  $\langle \{ \rm{approx} \} 1.5 \{ \{ \rm{m} \} \}_{\{ \{ \rm{s} \} \}}^{\{ -1 \}} \rangle$  is reached and maintained during all the measurements.

## Extended Data Fig. 5 Spin ensemble simulations.

**a**, Measured (blue dots) and simulated (orange solid line) electromagnetic field amplitude at the output of the spin cavity as a function of the time  $T$  from the  $\pi/2$  pulse of an echo sequence. The echo appears as a slight increase of the field amplitude at twice the separation between the  $\pi/2$  and  $\pi$  pulse (inset). The spin spectral density  $\langle \{ \rho \}_{\{ \{ \rm{spin} \} \}} \rangle$  is the only free parameter of the simulation, accordance with the experimental data is achieved for  $\langle \{ \rho \}_{\{ \{ \rm{s} \} \}_{\{ \{ \rm{p} \} \}} \{ \{ \rm{i} \} \}} \rangle$   
 $\langle \{ \rm{n} \} \rangle = 14.6 \{ \{ \rm{s} \} \}_{\{ \{ \rm{p} \} \}} \{ \{ \rm{i} \} \} \{ \{ \rm{n} \} \}_{\{ \{ \rm{s} \} \}_{\{ \{ \rm{k} \} \}}}$   
 $\{ \{ \rm{H} \} \}_{\{ \{ \rm{z} \} \}}^{\{ -1 \}}$ ). **b**, Simulated time evolution of  $\langle \{ S \}_{\{ \{ \rm{z} \} \}} \rangle$  during an echo sequence, using the same pulse parameters as in the experiment. **c**, Simulation of the time evolution of  $\langle \{ \langle \{ S \}_{\{ \{ \rm{z} \} \}} \rangle \rangle$  during a  $\pi$ -pulse with the same parameters of the experiment of photo-detected incoherent relaxation. A spin density  $\langle \{ \rho \}_{\{ \{ \rm{s} \} \}_{\{ \{ \rm{p} \} \}} \{ \{ \rm{i} \} \}} \rangle$   
 $\langle \{ \rm{n} \} \rangle = 12 \{ \{ \rm{s} \} \}_{\{ \{ \rm{p} \} \}} \{ \{ \rm{i} \} \}_{\{ \{ \rm{n} \} \}_{\{ \{ \rm{s} \} \}_{\{ \{ \rm{k} \} \}}}$   
 $\{ \{ \rm{H} \} \}_{\{ \{ \rm{z} \} \}}^{\{ -1 \}}$ ) was adjusted so that the ratio between the variation of  $\langle \{ \langle \{ S \}_{\{ \{ \rm{z} \} \}} \rangle \rangle$  at echo time (**b**) and upon the  $\pi$  pulse excitation (**c**) reproduces the experimental ratio  $\langle \{ C \}_{\{ \{ \rm{spin} \} \}} \rangle / (\langle \{ \eta \}_{\{ \{ \rm{duty} \} \}} \{ \{ \rm{c} \}_{\{ \{ \rm{e} \} \}} \rangle)$ .

## Supplementary information

### Peer Review File

# Rights and permissions

## [Reprints and Permissions](#)

## About this article

### Cite this article

Albertinale, E., Balembois, L., Billaud, E. *et al.* Detecting spins by their fluorescence with a microwave photon counter. *Nature* **600**, 434–438 (2021). <https://doi.org/10.1038/s41586-021-04076-z>

- Received: 19 January 2021
- Accepted: 30 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04076-z>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 15 December 2021](#)

# Fractional Chern insulators in magic-angle twisted bilayer graphene

- [Yonglong Xie](#) [ORCID: orcid.org/0000-0001-7656-6324](#)<sup>1,2</sup> natl,
- [Andrew T. Pierce](#)<sup>1</sup> natl,
- [Jeong Min Park](#) [ORCID: orcid.org/0000-0001-8905-3640](#)<sup>2</sup> natl,
- [Daniel E. Parker](#)<sup>1</sup>,
- [Eslam Khalaf](#)<sup>1</sup>,
- [Patrick Ledwith](#)<sup>1</sup>,
- [Yuan Cao](#) [ORCID: orcid.org/0000-0002-9000-4501](#)<sup>2</sup>,
- [Seung Hwan Lee](#)<sup>1</sup>,
- [Shaowen Chen](#)<sup>1</sup>,
- [Patrick R. Forrester](#) [ORCID: orcid.org/0000-0001-8882-9048](#)<sup>1</sup>,
- [Kenji Watanabe](#) [ORCID: orcid.org/0000-0003-3701-8119](#)<sup>3</sup>,
- [Takashi Taniguchi](#) [ORCID: orcid.org/0000-0002-1467-3105](#)<sup>4</sup>,
- [Ashvin Vishwanath](#)<sup>1</sup>,
- [Pablo Jarillo-Herrero](#) [ORCID: orcid.org/0000-0001-8217-8213](#)<sup>2</sup> &
- [Amir Yacoby](#) <sup>1</sup>

*Nature* volume 600, pages 439–443 (2021)

- 2722 Accesses
- 78 Altmetric
- [Metrics details](#)

## Subjects

- [Electronic properties and materials](#)
- [Topological matter](#)

## Abstract

Fractional Chern insulators (FCIs) are lattice analogues of fractional quantum Hall states that may provide a new avenue towards manipulating non-Abelian excitations. Early theoretical studies<sup>1,2,3,4,5,6,7</sup> have predicted their existence in systems with flat Chern bands and highlighted the critical role of a particular quantum geometry. However, FCI states have been observed only in Bernal-stacked bilayer graphene (BLG) aligned with hexagonal boron nitride (hBN)<sup>8</sup>, in which a very large magnetic field is responsible for the existence of the Chern bands, precluding the realization of FCIs at zero field. By contrast, magic-angle twisted BLG<sup>9,10,11,12</sup> supports flat Chern bands at zero magnetic field<sup>13,14,15,16,17</sup>, and therefore offers a promising route towards stabilizing zero-field FCIs. Here we report the observation of eight FCI states at low magnetic field in magic-angle twisted BLG enabled by high-resolution local compressibility measurements. The first of these states emerge at 5 T, and their appearance is accompanied by the simultaneous disappearance of nearby topologically trivial charge density wave states. We demonstrate that, unlike the case of the BLG/hBN platform, the principal role of the weak magnetic field is merely to redistribute the Berry curvature of the native Chern bands and thereby realize a quantum geometry favourable for the emergence of FCIs. Our findings strongly suggest that FCIs may be realized at zero magnetic field and pave the way for the exploration and manipulation of anyonic excitations in flat moiré Chern bands.

[Download PDF](#)

## Main

The search for novel material systems exhibiting topological properties holds promise for the next generation of electronics. For example, band-structure engineering guided by theoretical predictions has enabled the realization of integer quantized Hall states at zero magnetic field<sup>18,19,20</sup>, enabling new directions in spintronics and topological quantum computing. Likewise, extensive efforts have been directed towards engineering FCIs—lattice analogues of fractional quantum Hall (FQH) states—in part because of their potential to manifest high-temperature topological order and to host non-Abelian excitations at zero magnetic field. However, despite a large body of theoretical work<sup>1,2,3,4,5,6,7</sup>, FCI states have proved exceptionally difficult to stabilize experimentally, as they require not only non-dispersive Chern bands, but also a particular quantum band geometry including a flat Berry curvature distribution. To date, FCI states have been observed only in Hofstadter bands of a BLG heterostructure aligned with hBN at very large ( $\sim 30$  T) magnetic fields<sup>8</sup>. A key disadvantage of this platform is that its band topology fundamentally originates from the presence of the magnetic field, thus precluding the realization of FCIs in the zero-field limit.

By contrast, moiré superlattices with native topological bands<sup>13,14,15,16,17</sup> provide a promising avenue to search for FCIs at zero magnetic field. In particular, the recent discovery of correlated Chern insulators (ChIs) in magic-angle twisted BLG (MATBG) down to zero field confirms the presence of intrinsic flat Chern bands<sup>20,21,22,23,24,25,26,27,28,29</sup> and thus raises the possibility of realizing FCIs in this system. Indeed, recent analytical considerations<sup>30</sup> and numerical calculations<sup>31,32,33</sup> have predicted FCI ground states in MATBG aligned with hBN. Importantly, these works also show the close competition between FCIs and other correlated phases such as charge density waves (CDWs), and highlight the importance of Berry curvature distribution homogeneity and the quantum metric in stabilizing FCIs in MATBG. Here we report the observation of eight FCI states at fractional fillings of the Chern bands in MATBG. The first of these states appears at 5 T in the range  $3 < v < 4$ , where the system is well described by an isolated Chern band. We show that these FCI states result from the intrinsic band topology of MATBG and are stabilized by weak magnetic fields that create favourable quantum geometric conditions for their emergence. The FCIs observed beyond this range, where the parent Chern states possibly reacquire their multicomponent character, are more complex, probably owing to the interplay between multiple degrees of freedom, and demonstrate the potential of MATBG for exploring novel emergent topological order.

## Correlated phases at fractional fillings

To search for such topological states, we perform local electronic compressibility measurements on an MATBG device aligned with the hBN with a twist angle of  $\sim 1.06^\circ$  (see [Methods](#)) using a scanning single-electron transistor (SET). Our measurements of the inverse compressibility  $d\mu/dn$  as a function of perpendicular magnetic field  $B$  and moiré band filling factor  $v$  reveal a large number of linearly dispersing incompressible states (Fig. [1a,b](#)) that can be classified by a pair of quantum numbers  $(t, s)$  satisfying the Diophantine equation  $\lfloor \nu = t\varphi / \{ \varphi \}_0 + s \rfloor$ , where  $v$  is the filling factor at which the incompressible peak occurs,  $\phi$  is the magnetic flux per moiré unit cell, and  $\phi_0$  is the magnetic flux quantum. We observe in total five distinct classes of incompressible states. First, incompressible features with  $t=0$  and integer  $s \neq 0$  correspond to trivial correlated insulators (green line in Fig. [1b](#)). Second, features with integer  $t \neq 0$  and integer  $s$  correspond to integer quantum Hall states or ChIs (black lines in Fig. [1b](#)), some of which have been identified as translation symmetry (TS)-broken states resulting from unit-cell doubling. The observed ChIs at zero field are the parent states essential for realizing more complex topological states in the zero-field limit. Finally, we observe three classes of gapped states with fractional  $t$  and/or  $s$ , which we identify as CDWs ( $t=0$  and fractional  $s$ ), symmetry-broken ChIs (SBCIs—integer  $t \neq 0$  and fractional  $s$ ) and FCIs (fractional  $t$  and fractional  $s$ ).

**Fig. 1: Incompressible states with fractional quantum numbers in MATBG.**

 figure1



**a**, Local inverse compressibility  $d\mu/dn$  measured as a function of magnetic field  $B$  and electrons per moiré unit cell  $\nu$ . **b**, Wannier diagram identifying the incompressible peaks present in **a**. Black lines correspond to ChIs and integer quantum Hall (IQH) states; green lines correspond to correlated insulators (CIs) emanating with nonzero integer  $s$  and  $t = 0$ ; blue lines correspond to CDWs with integer  $t = 0$  and fractional  $s$ ; yellow lines correspond to SBCIs with nonzero integer  $t$  and fractional  $s$ ; and orange lines correspond to FCIs with fractional  $t$  and fractional  $s$ . Grey shaded regions correspond to the gaps to the remote bands.

[Source data](#)

To demonstrate that our system provides the topological bands and strong correlations essential for the realization of FCIs, we focus on the range of filling factors near  $\nu = 3$ , as in this density range the band structure can be best approximated by isolated Chern

bands. Figure 2a shows a measurement of inverse compressibility as a function of magnetic field for  $2.5 < \nu < 4$  for  $B < 3$  T. In addition to the insulators emanating from  $\nu = 3$ , we discover three new incompressible states that are stable down to zero magnetic field: the two non-dispersive states  $(0, 7/2)$  and  $(0, 11/3)$ , which we classify as trivial CDWs, and the SBCI state  $(1, 8/3)$  (Fig. 2b). The fractional values of  $s$  associated with these states strongly suggest that electron–electron interactions spontaneously break the TS of the underlying moiré superlattice. In fact, a previous study<sup>29</sup> has shown that the appearance of a portion of the ChIs is probably a consequence of TS breaking via doubling of the unit cell. In this scenario, the Hartree potential favours filling states near the centre of the mini-Brillouin zone, which in MATBG is also the region where the Berry curvature is highly concentrated (yellow trace in Fig. 2c). Consequently, the system may favour forming one band that retains the original Berry curvature and therefore has  $C = \pm 1$ , along with a new  $C = 0$  band (Fig. 2d). Under this assumption, filling three of the four  $C = 0$  bands generated by unit-cell doubling yields the  $(0, 7/2)$  state (Fig. 2f). Similarly, tripling the unit cell allows one  $C = \pm 1$  band to give rise to a  $C = \pm 1$  band accompanied by two  $C = 0$  bands (Fig. 2e). Sequentially filling the 12 reconstructed bands produces both the  $(1, 8/3)$  and  $(0, 11/3)$  states (Fig. 2f,g). Together, the observation of CDW and SBCI states at zero field establishes the presence of both intrinsic band topology and strong electron–electron interactions, and highlights the critical role of the non-uniform Berry curvature in stabilizing these two classes of states.

**Fig. 2: Density wave states at low magnetic field for  $2.5 < \nu < 4$ .**



**a**, Local inverse compressibility  $d\mu/dn$  between  $v = 2.5$  and  $4$  for  $B = 0$  T to 3 T. **b**, Wannier diagram corresponding to the states observed in **a** coloured according to the classification used in Fig. 1b. **c**, Band energy (purple) and Berry curvature (yellow) along a path through the  $\Gamma$  point in the first mini-Brillouin zone. Zero momentum corresponds to the  $\Gamma$  point. **d**, Band structure in the case of unit-cell (UC) doubling resulting in a  $C = \pm 1$  band accompanied by a new  $C = 0$  band. **e**, Band structure in the case of unit-cell tripling resulting in a  $C = \pm 1$  band accompanied by two new  $C = 0$  bands. **f–h**, Band fillings in the case of unit-cell doubling (**f**) and unit-cell tripling (**g**, **h**) needed to produce the density wave states observed in **a**.

### Source data

Remarkably, on increasing the magnetic field to 5 T, we observe a different family of robust incompressible states that are parametrized by fractional values of both  $t$  and  $s$  (Fig. 3a,b), characteristic of FCIs. These states,  $(2/3, 10/3)$  and  $(1/3, 11/3)$ , persist up to at least 11 T, and can be interpreted as lattice analogues of  $v_c = 1/3$  and  $2/3$  FQH states from the final  $C = -1$  band populated on electron-doping the  $(1, 3)$  ChI, where  $v_c$  is the filling factor of the partially filled Chern band (Fig. 3c). As these states do not require breaking of the TS of the moiré superlattice, they are referred to below as symmetry-preserving FCIs. These two states are expected to exhibit fractional quantized Hall conductance according to the Streda formula and hence support quasiparticle excitations with fractional charge  $e/3$  (ref. 34). Integrating  $d\mu/dn$  with respect to the electron density allows us to directly extract the steps in chemical potential  $(\Delta \mu)$  associated with each of the observed CDW and FCI states (Fig. 3d). As the chemical potential is defined with respect to electrons,  $(\Delta \mu)$  must be multiplied by the ratio of the quasiparticle charge to the electron charge, yielding energy gaps of about  $50 \pm 20 \mu\text{eV}$  ( $\sim 0.6$  K) for both FCI states, roughly in agreement with the estimate of  $0.01U$  from a recent exact diagonalization study<sup>33</sup>, where  $U$  is the strength of Coulomb interaction. The same study also argues that, because the spin polarization of the valley-polarized Chern band is unknown, the FCI states can be either isospin-polarized Laughlin states or multicomponent states depending on the detailed quantum geometric properties of the system. While our measurements are not capable of directly distinguishing between single and multicomponent ground states, we note that the gaps associated with both FCIs are much smaller than the spin Zeeman energy scale  $E_Z = g\mu_B B$  (assuming  $g = 2$ ), where  $\mu_B$  is the Bohr magneton, and depend very weakly on  $B$ , suggesting that the charged excitations of both states probably do not require a spin flip. The sudden appearance of the FCIs and disappearance of the CDWs indicates close competition between these two phases, with the magnetic field driving the transition yet leaving the band topology unaltered.

**Fig. 3: FCIs in a weak magnetic field.**

---

 **figure3**

---

**a**, Local inverse compressibility  $d\mu/dn$  between  $\nu = 3$  and 4 for  $B = 3$  T to 11 T. **b**, Wannier diagram corresponding to the states observed in **a** coloured according to the classification used in Fig. 1b. Light blue and orange lines denote the CDWs and FCIs, respectively. The grey shaded region marks the energy gap to the remote band. **c**, Depiction of band fillings that lead to the  $(t, s) = (2/3, 10/3)$  FCI observed in **a**, which corresponds to a  $\nu_c = 1/3$  FCI state from the final  $C = -1$  band populated on electron-doping the  $(1, 3)$  ChI. **d**, Chemical potential steps  $\Delta\mu$  associated with the CDW and FCI states observed in **a** obtained by integrating the inverse compressibility  $d\mu/dn$ . The error bars reflect the standard deviation obtained from fitting to  $\mu(n)$ . **e**, Calculated average Berry curvature deviation  $\langle|\sigma_c|_{\text{left}}(F)\rangle$  from the continuum model as a function of  $w_0/w_1$ . The grey shaded region corresponds to  $w_0/w_1 = 0.65$  to 0.75, the range in which the transition from FCI to CDW occurs. This  $w_0/w_1$  range allows us to estimate the range of values  $\langle|\sigma_c|_{\text{left}}(F)\rangle = 1.4$  to 2.2 below which the FCI is favourable. **f**, Calculated average Berry curvature deviation  $\langle|\sigma_c|_{\text{left}}(F)\rangle$  as a function of magnetic field for  $w_0/w_1 = 0.8$ . **g**, Phase diagram constructed as a function of  $w_0/w_1$  and magnetic field in units of  $\langle|\varphi_0|\rangle$ . The white lines indicate the contours  $\langle|\sigma_c|_{\text{left}}(F)=1.4\rangle$  and  $\langle|\sigma_c|_{\text{left}}(F)=2.2\rangle$  that define the region where the phase boundary between the FCI and CDW ground states is expected.

[Source data](#)

## Quantum geometry of MATBG

To understand the transition from a CDW-dominated to an FCI-dominated regime, we begin by observing that these two classes of ground states place very different

constraints on the quantum geometric properties of the underlying band structure of MATBG. For the CDW ground states to emerge, the Berry curvature of the flat bands must be strongly concentrated near the centre of the mini-Brillouin zone to take advantage of the Hartree potential<sup>29</sup>. However, bands with sufficiently nonuniform Berry curvature are known to disfavour FCI ground states<sup>30,31,32,33</sup>. The observed transition therefore suggests that the applied magnetic field in the experiment serves primarily to reduce the intrinsic Berry curvature inhomogeneity within the partially filled Chern band, unlike in the hBN/BLG system where the applied field is needed to produce Chern bands in the first place. To estimate the amount of Berry curvature inhomogeneity the FCI ground states can tolerate, we note that exact diagonalization studies<sup>31,32,33</sup> indicate a transition between CDW and FCI ground states as a function of  $w_0/w_1$ , where  $w_0$  and  $w_1$  are the interlayer tunnelling matrix elements at the AA-stacked and the AB-stacked regions, respectively. This ratio is known to strongly alter the Berry curvature distribution within the flat bands of MATBG. According to these works, the transition occurs near  $w_0/w_1 \approx 0.7$ , as has been confirmed by a recent density matrix renormalization group study (D.E.Parker et al., manuscript in preparation). The ground-state dependence on  $w_0/w_1$  therefore gives a means of parametrizing the dependence of the FCI ground states on the Berry curvature inhomogeneity, which we characterize using the quantity  $\langle(\sigma_{\text{left}(F)}\rangle)$ , the mean standard deviation of the Berry curvature over the mini-Brillouin zone. We estimate an upper bound on the allowable Berry curvature inhomogeneity  $\langle(\sigma_{\text{left}(F)}\rangle)$  to be in the range of 1.4 to 2.2, depending on the model parameters, below which the quantum geometry of the system is favourable for the emergence of FCIs (Fig. 3e). For simplicity, we choose  $\langle(\sigma_{\text{left}(F)}\rangle) = 1.8$  for the discussion below. For realistic MATBG samples,  $w_0/w_1$  is estimated to be around 0.8 (refs. <sup>25,35,36</sup>), yielding large values of  $\sigma(F) \sim 3$ , consistent with our observation of CDW states at zero field. Thus, the absence of FCI ground states at zero magnetic field in our device can be understood to result from the large values of  $\langle(\sigma_{\text{left}(F)}\rangle)$  present in MATBG.

Having established the critical role of  $\langle(\sigma_{\text{left}(F)}\rangle)$  in determining the many-body ground state, we now examine its evolution as a function of magnetic field by analysing the Hofstadter spectrum of the continuum model of MATBG aligned with hBN (see [Methods](#)). We find that increasing the magnetic field reduces  $\langle(\sigma_{\text{left}(F)}\rangle)$  monotonically (Fig. 3f), with  $\langle(\sigma_{\text{left}(F)}\rangle)$  vanishing as  $\langle(\varphi/\{\varphi\}_0)\rangle$  to 1. To estimate the value of magnetic field at which the FCI ground state becomes favourable, we identify the magnetic field,  $B_c$ , at which  $\langle(\sigma_{\text{left}(F)}\rangle)$  falls below the critical value  $\langle(\sigma_{\text{left}(F)}\rangle) \approx 1.8$ . For realistic values of  $w_0/w_1 \sim 0.8$ , our calculations find that  $\langle(\sigma_{\text{left}(F)}\rangle)$  is reduced below  $\langle(\sigma_{\text{left}(F)}\rangle)$  starting at  $\langle(\varphi/\{\varphi\}_0)\rangle \sim 1/5$  or  $B_c \sim 5.4$  T, in good

agreement with the magnetic field at which the  $(2/3, 10/3)$  and  $(1/3, 11/3)$  states appear experimentally. We emphasize that due to the sharp decrease of  $\langle\langle\sigma_z\rangle\rangle$  with field, the critical field  $B_c$  is not sensitive to the precise choice of  $\langle\langle\sigma_z\rangle\rangle$ . Combining the bound  $\langle\langle\sigma_z\rangle\rangle$  estimated from many-body ground state analyses<sup>[31,32,33](#) (D.E.P., manuscript in preparation) at  $B = 0$  with our calculations of  $\langle\langle\sigma_z\rangle\rangle$  as a function of  $B$  and  $w_0/w_1$  allows us to sketch a phase diagram at  $v_c = 1/3$  (Fig. [3g](#)). Our calculations also demonstrate that the FCI is adiabatically connected to the FQH state at  $\langle\langle\varphi/\varphi_0\rangle\rangle = 1$ , where the band geometry reduces to that of the lowest Landau level. However, unlike the case of the usual FQH states or of FCIs occurring within partially filled Hofstadter bands in a BLG/hBN heterostructure, in which the Berry curvature is supplied by the Landau levels or Chern bands that form in a magnetic field, the FCIs observed here fundamentally stem from zero-field ChI parent states, and the only role of the magnetic field is to flatten the Berry curvature. Therefore, only a weak magnetic field of less than 20% of a magnetic flux quantum per moiré unit cell is required to stabilize the FCIs by reducing  $\langle\langle\sigma_z\rangle\rangle$  below  $\langle\langle\sigma_z\rangle\rangle_{\rm c}$ .</sup>

## FCIs away from $3 < v < 4$

Outside the density range  $3 < v < 4$ , the system recovers additional degrees of freedom and thus permits more possible competing ground states at fractional fillings. In particular, we observe six additional FCIs—along with numerous SBCIs with denominators of  $s$  as large as 10 (Extended Data Fig. [1](#))—at slightly higher values of magnetic field, particularly on the hole side (Fig. [4](#)), most of which show values of  $\langle\langle\Delta\mu\rangle\rangle$  comparable to those of their counterparts near  $v = 3$  (Extended Data Fig. [2](#)). We emphasize that our measurements unambiguously identify these states as FCIs purely on the basis of the Streda formula regardless of their exact nature and origin, on which we speculate below. As in the case of the FCIs observed for  $3 < v < 4$ , several of these additional states probably correspond to symmetry-preserving FCIs (Fig. [4a,b](#)). For example, we interpret the state  $(-4/3, -5/3)$  as arising from a  $v_c = 1/3$  FCI formed within the  $C = -1$  band populated on electron-doping the  $(-1, -2)$  ChI, similar to the  $(2/3, 10/3)$  and  $(1/3, 11/3)$  states described above. In addition, unlike the aforementioned states, the observed  $(-8/5, 11/10)$  and  $(-7/3, 2/9)$  states (Fig. [4c,d](#)) have  $s$  with denominator a multiple of that of  $t$ , rather than being equal, suggesting that each unit cell binds only a fraction of an electron charge and that the states therefore correspond to symmetry-broken FCIs. Specifically, the  $(-8/5, 11/10)$  state can result from doping the  $(-2, 1)$  ChI with a  $v_c = 2/5$  FCI that quadruples the unit cell and thus contributes to a change in  $s$  of only  $1/10$ . This interpretation is further supported by the fact that the  $(-2, 1)$  ChI is a state that breaks TS, and thus might naturally be expected to also support symmetry-broken FCIs. We note that such symmetry-broken FCI states

have no analogue in the FQH system. Most intriguingly, we also find evidence of FCI states with coprime denominators of  $t$  and  $s$  that as a result cannot be described as either symmetry-preserving or symmetry-breaking FCIs. The emergence of these exotic many-body ground states may originate from complex interplay between spin, valley and spatial symmetry.

**Fig. 4: Additional FCIs at higher magnetic field.**



**a–d**, Measurements of  $d\mu/dn$  ( $\times 10^{-11}$  mV cm $^{-2}$ ) in various density ranges between 6 and 12 T showing additional FCIs and SBCIs. **e–h**, Wannier diagrams corresponding to the states observed in **a–d** coloured according to the classification used in Fig. 1b. Green, yellow and orange lines denote the correlated insulators, SBCIs and FCIs, respectively.

#### Source data

The observation of FCIs in MATBG reported here leaves open many theoretical and experimental questions. An interesting and straightforward direction is to identify the quasiparticle charge associated with these FCI states, especially those that have no analogues in the FQH system. The competition between FCIs and nearby CDWs and SBCIs may provide a new setting for the study of quantum phase transitions.

Importantly, our work establishes the applied magnetic field as a novel tuning knob for the Berry curvature distribution, and indicates close proximity to zero-field FCIs in the flat bands of MATBG. Thus, a pressing experimental task is to develop means of reducing  $w_0/w_1$  in MATBG and to explore alternative platforms beyond MATBG that suffer less from Berry curvature inhomogeneity, which would enable the realization of FCIs at zero magnetic field and offer new opportunities for the creation of next-generation topological quantum devices.

# Methods

## Sample preparation

The MATBG device used in this study was fabricated using the ‘tear-and-stack’ technique described in refs. [37,38](#), and is the same as the one in ref. [29](#). Briefly, the monolayer graphene and hBN flakes were first exfoliated on SiO<sub>2</sub>/Si substrates and subsequently screened with optical microscopy and atomic force microscopy. We use a PC/PDMS stamp on a glass slide to sequentially pick up the flakes. The resulting stack is released on the pre-stacked hBN-on-Pd/Au back gate. The device geometry was defined by electron-beam lithography and reactive ion etching. Cr/Au electrical contacts to MATBG were made by the standard edge-contact method.

## Compressibility measurements

All compressibility measurements were made in a <sup>3</sup>He cryostat. The SET tips were fabricated using a procedure described elsewhere [39](#). Compressibility measurements were performed using a.c. and d.c. protocols similar to those described in refs. [29,39](#). a.c. excitations of 40 mV at 97.17 Hz and 1.5 mV at 107.17 Hz were applied to the back gate and to MATBG, respectively. The tip was held approximately 100 nm above the MATBG. A d.c. feedback loop was used to hold the phase of the SET’s Coulomb blockade signal fixed, which results in a direct d.c. measurement of the chemical potential. Compared to ref. [29](#), a longer integration and voltage-ramping time and a slightly smaller tip-sample distance were used to further improve the signal-to-noise ratio.

## Procedure of determining the slope and the intercept for the fractional states

We determine the quantum numbers ( $t, s$ ) of the incompressible states by first identifying the peaks associated with each state and performing a linear fit to obtain their slope and intercept. To more accurately confirm the fractional values of  $t$  and  $s$  and mitigate the error due to effects of quantum capacitance [8](#), we use the fitted slope and intercept of the nearby Chern and correlated insulators to obtain local estimates of  $t$  and  $s$ . On the basis of the converted values of  $t$  and  $s$ , we assign the corresponding fractions for  $t$  and  $s$  by identifying those with the smallest denominator possible (up to 10) within the 95% confidence interval and favour the fractions of  $t$  and  $s$  that share the same denominators (Extended Data Fig. [3](#)).

## TS breaking

We briefly summarize the origin of TS breaking in MATBG and its consequence on the topological Chern structure reported in ref. [29](#). It has been shown that the strong Hartree potential of the flat bands of MATBG favours populating states at the centre (corner) of MATBG’s mini-Brillouin zone for the electron-doped (hole-doped) side (see Fig. [2c](#) for the electron-doped case). As a result, the system can lower its energy by enlarging the moiré unit cell (doubling, tripling, quadrupling and so on), thereby breaking TS to form an insulator. The immediate consequence of enlarging the moiré unit cell is to fold the original bands, which results in a new set of  $N$  reconstructed bands, where  $N$  is the factor by which the unit cell is enlarged. For example, doubling (tripling) the unit cell leads to two (three) reconstructed bands (Fig. [2d,e](#)) per flavour. As the Berry curvature is highly concentrated at the centre of the mini-Brillouin zone (Fig. [2c](#)), the lowest (highest) band retains the Chern number of the original band for the electron-doped (hole-doped) side, while the rest of the reconstructed bands carry zero Chern number. This picture captures the unconventional sequence of ChIs as well as the zero-field SBCIs and CDWs we observe in the present work.

## Influence of twist angle inhomogeneity on FCIs

We have examined the twist angle inhomogeneity over a distance of approximately  $1.6\text{ }\mu\text{m}$ , and we find the local twist angle to vary by approximately  $0.005^\circ$  (Extended Data Fig. [4a](#)). Extended Data Fig. [4b](#) shows the compressibility measurements between  $\nu = 3$  and 4 taken at 9 T, where the incompressible peaks associated with (0, 3) and (1, 3) shift in density owing to the changes in the local twist angle. Similarly, the incompressible features that appear in the density range where the (2/3, 10/3) and (1/3, 11/3) FCI states are expected to occur—indicated by the blue and black arrows, respectively—also display shifts in density. To unambiguously identify the nature of these incompressible peaks, we have examined their magnetic field dependence at three different locations (Extended Data Fig. [4c](#)). We find that, at location 2, where the variation of twist angle is small, the evolution of the two incompressible peaks indeed follows the Diophantine equation with  $t = 2/3$  and  $t = 1/3$ , demonstrating the robust reproducibility of the observation of the FCIs. At location 1, the peaks associated with  $t = 2/3$  and  $t = 1/3$  are also present over most of the field range, but the appearance of additional incompressible peaks suggests that location 1 is near a region where local disorder different from twist angle inhomogeneity is present. However, at location 3 where the twist angle varies rapidly, the left incompressible peak deviates from the expected trajectory for a  $t = 2/3$  FCI and may be better described by a  $C = 1$  SBCI emanating from  $s = 13/4$ . Overall, these measurements demonstrate the robustness of the FCI ground states and highlight the critical role of twist angle homogeneity in stabilizing the FCIs, lending further support to the idea that controlling local microscopic parameters may provide a pathway for tuning transitions between CDWs/SBCIs and FCIs.

## Hofstadter spectrum

We model the system using the Bistritzer–MacDonald model<sup>9</sup> with a twist angle of  $\theta = 1.06^\circ$ , and account for the gap at charge neutrality observed in the experiment by including a sublattice splitting of 30 meV. The interlayer tunnelling parameter  $w_1$  is set to 110 meV. In the range  $3 < \nu < 4$ , the system can be approximated by a single Chern band, and we thus consider a single fermion species in our calculation. We obtain the Hofstadter spectrum following refs. [40,41,42](#), which is shown in Extended Data Fig. 5. As dictated by the Streda formula, the top  $C = -1$  band—the parent state of the FCI—is separated by a gap. Complete details of the model are given in the [Supplementary Information](#).

## Quantum geometry

The stability of FCI is closely related to the quantum geometry of the MATBG band structure. One key figure of merit is the standard deviation of the Berry curvature distribution over the Brillouin zone. In the presence of a magnetic flux  $\varphi = \frac{p}{q} \varphi_0$ , there are  $2q$  bands in the Hofstadter spectrum, where  $(q - \lfloor \frac{C}{2} \rfloor)$  bands are filled. Here we present a natural multi-band generalization for the standard deviation of the Berry curvature, which is continuous, gauge invariant and reduces to the expected values at  $\varphi = 0$ . At a given magnetic flux  $\varphi$ , let  $P_k = \sum_{a=1}^N |u_a(k)\rangle\langle u_a(k)|$  be the projector to the top  $C = -1$  band. The  $U(N)$  non-Abelian Berry curvature is defined as  $F_{ab} = \langle \partial_x u_a | \partial_y u_b \rangle - \langle \partial_y u_a | \partial_x u_b \rangle$ , where  $\partial_\mu = \frac{\partial}{\partial k_\mu}$  and  $A = \frac{1}{2\pi} \varphi / N$  is the area of the magnetic Brillouin zone. We note that the non-standard normalization  $NA$  is necessary for gauge-invariant quantities to be continuous functions of magnetic field. A semi-analytic formula for  $F$ , which is both numerically stable and accounts for the intrinsic geometry, is derived in the [Supplementary Information](#). To evaluate expectation values of the Berry curvature distribution, we define the trace operator  $\text{tr}[O] = \frac{1}{N} \sum_{k=1}^N \langle P_k | O | P_k \rangle$  so that  $\text{tr}[I] = 1$ , where  $I$  is the identity operator. The Chern number  $C = \text{tr}[F]$  is the mean of the distribution, up to  $2\pi$ . The standard deviation of the Berry curvature is then defined as  $\sigma_F = \sqrt{\text{tr}[F^2] - C^2}$ .

## Data availability

All data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

## Code availability

The code that supports the findings of this study is available from the corresponding authors upon reasonable request.

## References

1. 1.

Sheng, D. N., Gu, Z.-C., Sun, K. & Sheng, L. Fractional quantum Hall effect in the absence of Landau levels. *Nat. Commun.* **2**, 389 (2011).

2. 2.

Neupert, T., Santos, L., Chamon, C. & Mudry, C. Fractional quantum Hall states at zero magnetic field. *Phys. Rev. Lett.* **106**, 236804 (2011).

3. 3.

Regnault, N. & Bernevig, B. A. Fractional Chern insulator. *Phys. Rev. X* **1**, 021014 (2011).

4. 4.

Tang, E., Mei, J.-W. & Wen, X.-G. High-temperature fractional quantum Hall states. *Phys. Rev. Lett.* **106**, 236802 (2011).

5. 5.

Qi, X.-L. Generic wave-function description of fractional Quantum anomalous Hall states and fractional topological insulators. *Phys. Rev. Lett.* **107**, 126803 (2011).

6. 6.

Parameswaran, S. A., Roy, R. & Sondhi, S. L. Fractional quantum Hall physics in topological flat bands. *C. R. Phys.* **14**, 816–839 (2013).

7. 7.

Bergholtz, E. J. & Liu, Z. Topological flat band models and fractional Chern insulators. *Int. J. Mod. Phys. B* **27**, 1330017 (2013).

8. 8.

Spanton, E. M. et al. Observation of fractional Chern insulators in a van der Waals heterostructure. *Science* **360**, 62–66 (2018).

9. 9.

Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci. USA* **108**, 12233–12237 (2011).

10. 10.

Suárez Morell, E., Correa, J. D., Vargas, P., Pacheco, M. & Barticevic, Z. Flat bands in slightly twisted bilayer graphene: tight-binding calculations. *Phys. Rev. B* **82**, 121407 (2010).

11. 11.

Cao, Y. et al. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018).

12. 12.

Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).

13. 13.

Po, H. C., Zou, L., Vishwanath, A. & Senthil, T. Origin of Mott insulating behavior and superconductivity in twisted bilayer graphene. *Phys. Rev. X* **8**, 031089 (2018).

14. 14.

Song, Z. et al. All magic angles in twisted bilayer graphene are topological. *Phys. Rev. Lett.* **123**, 036401 (2019).

15. 15.

Ahn, J., Park, S. & Yang, B.-J. Failure of Nielsen-Ninomiya theorem and fragile topology in two-dimensional systems with space-time inversion symmetry:

application to twisted bilayer graphene at magic angle. *Phys. Rev. X* **9**, 021013 (2019).

16. 16.

Po, H. C., Zou, L., Senthil, T. & Vishwanath, A. Faithful tight-binding models and fragile topology of magic-angle bilayer graphene. *Phys. Rev. B* **99**, 195455 (2019).

17. 17.

Liu, J., Liu, J. & Dai, X. Pseudo Landau level representation of twisted bilayer graphene: band topology and implications on the correlated insulating phase. *Phys. Rev. B* **99**, 155415 (2019).

18. 18.

Chang, C.-Z. et al. Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).

19. 19.

Deng, Y. et al. Quantum anomalous Hall effect in intrinsic magnetic topological insulator MnBi<sub>2</sub>Te<sub>4</sub>. *Science* **367**, 895–900 (2020).

20. 20.

Serlin, M. et al. Intrinsic quantized anomalous Hall effect in a moiré heterostructure. *Science* **367**, 900–903 (2020).

21. 21.

Sharpe, A. L. et al. Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene. *Science* **365**, 605–608 (2019).

22. 22.

Nuckolls, K. P. et al. Strongly correlated Chern insulators in magic-angle twisted bilayer graphene. *Nature* <https://doi.org/10.1038/s41586-020-3028-8> (2020).

23. 23.

Wu, S., Zhang, Z., Watanabe, K., Taniguchi, T. & Andrei, E. Y. Chern insulators, van Hove singularities and topological flat bands in magic-angle twisted bilayer

graphene. *Nat. Mater.* **20**, 488–494 (2021).

24. 24.

Saito, Y. et al. Hofstadter subband ferromagnetism and symmetry-broken Chern insulators in twisted bilayer graphene. *Nat. Phys.* **17**, 478–481 (2021).

25. 25.

Das, I. et al. Symmetry-broken Chern insulators and Rashba-like Landau-level crossings in magic-angle bilayer graphene. *Nat. Phys.* <https://doi.org/10.1038/s41567-021-01186-3> (2021).

26. 26.

Choi, Y. et al. Correlation-driven topological phases in magic-angle twisted bilayer graphene. *Nature* **589**, 536–541 (2021).

27. 27.

Park, J. M., Cao, Y., Watanabe, K., Taniguchi, T. & Jarillo-Herrero, P. Flavour Hund's coupling, Chern gaps and charge diffusivity in moiré graphene. *Nature* **592**, 43–48 (2021).

28. 28.

Stepanov, P. et al. Competing zero-field Chern insulators in superconducting twisted bilayer graphene. Preprint at <https://arxiv.org/abs/2012.15126> (2020).

29. 29.

Pierce, A. T. et al. Unconventional sequence of correlated Chern insulators in magic-angle twisted bilayer graphene. *Nat. Phys.* **17**, 1210–1215 (2021).

30. 30.

Ledwith, P. J., Tarnopolsky, G., Khalaf, E. & Vishwanath, A. Fractional Chern insulator states in twisted bilayer graphene: an analytical approach. *Phys. Rev. Res.* **2**, 023237 (2020).

31. 31.

Abouelkomsan, A., Liu, Z. & Bergholtz, E. J. Particle-hole duality, emergent Fermi liquids, and fractional Chern insulators in moiré flatbands. *Phys. Rev. Lett.*

**124**, 106803 (2020).

32. 32.

Wilhelm, P., Lang, T. C. & Läuchli, A. M. Interplay of fractional Chern insulator and charge density wave phases in twisted bilayer graphene. *Phys. Rev. B* **103**, 125406 (2021).

33. 33.

Repellin, C. & Senthil, T. Chern bands of twisted bilayer graphene: fractional Chern insulators and spin phase transition. *Phys. Rev. Res.* **2**, 023238 (2020).

34. 34.

Laughlin, R. B. Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations. *Phys. Rev. Lett.* **50**, 1395–1398 (1983).

35. 35.

Carr, S., Fang, S., Zhu, Z. & Kaxiras, E. Exact continuum model for low-energy electronic states of twisted bilayer graphene. *Phys. Rev. Res.* **1**, 013001 (2019).

36. 36.

Xie, Y. et al. Spectroscopic signatures of many-body correlations in magic-angle twisted bilayer graphene. *Nature* **572**, 101–105 (2019).

37. 37.

Kim, K. et al. van der Waals heterostructures with high accuracy rotational alignment. *Nano Lett.* **16**, 1989–1995 (2016).

38. 38.

Cao, Y. et al. Superlattice-induced insulating states and valley-protected orbits in twisted bilayer graphene. *Phys. Rev. Lett.* **117**, 116804 (2016).

39. 39.

Martin, J., Feldman, B. E., Weitz, R. T., Allen, M. T. & Yacoby, A. Local compressibility measurements of correlated states in suspended bilayer graphene. *Phys. Rev. Lett.* **105**, 256806 (2010).

40. 40.

Bistritzer, R. & MacDonald, A. H. Moiré butterflies in twisted bilayer graphene. *Phys. Rev. B* **84**, 035440 (2011).

41. 41.

Hejazi, K., Liu, C. & Balents, L. Landau levels in twisted bilayer graphene and semiclassical orbits. *Phys. Rev. B* **100**, 035115 (2019).

42. 42.

Zhang, Y.-H., Po, H. C. & Senthil, T. Landau level degeneracy in twisted bilayer graphene: role of symmetry breaking. *Phys. Rev. B* **100**, 125104 (2019).

## Acknowledgements

We acknowledge discussions with A. Stern, Y. Sheffer and C. Repellin. This work was primarily supported by the US Department of Energy, Basic Energy Sciences Office, Division of Materials Sciences and Engineering under award DE-SC0001819 and the Gordon and Betty Moore Foundation through grant GBMF9468 to A.Y. Fabrication of samples was supported by the US Department of Energy, Basic Energy Sciences Office, Division of Materials Sciences and Engineering under award DE-SC0019300 and ARO grant number W911NF-14-1-0247. Help with transport measurements and data analysis were supported by the National Science Foundation (DMR-1809802), and the STC Center for Integrated Quantum Materials (NSF grant number DMR-1231319) (Y.C.). P.J.-H. acknowledges support from the Gordon and Betty Moore Foundation's EPiQS Initiative through grant GBMF9643. A.T.P. and P.L. acknowledge support from the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program. Y.X. and S.C. acknowledge partial support from the Harvard Quantum Initiative in Science and Engineering. A.T.P., Y.X. and A.Y. acknowledge support from the Harvard Quantum Initiative Seed Fund. A.V. was supported by a Simons Investigator award and by the Simons Collaboration on Ultra-Quantum Matter, which is a grant from the Simons Foundation (651440, A.V.). E.K. was supported by a Simons Investigator Fellowship, by NSF-DMR 1411343, and by the German National Academy of Sciences Leopoldina through a Leopoldina fellowship (grant LPDS 2018-02). P.R.F. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant number DGE 1745303. This research is funded in part by the Gordon and Betty Moore Foundation's EPiQS Initiative, Grant GBMF8683 to D.E.P. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan, grant number JPMXP0112101001, and JSPS KAKENHI grant number JP20H00354. This work was

performed, in part, at the Center for Nanoscale Systems, a member of the National Nanotechnology Infrastructure Network, which is supported by the NSF under award no. ECS-0335765. The Center for Nanoscale Systems is part of Harvard University.

## Author information

### Author notes

1. These authors contributed equally: Yonglong Xie, Andrew T. Pierce, Jeong Min Park

### Affiliations

1. Department of Physics, Harvard University, Cambridge, MA, USA

Yonglong Xie, Andrew T. Pierce, Daniel E. Parker, Eslam Khalaf, Patrick Ledwith, Seung Hwan Lee, Shaowen Chen, Patrick R. Forrester, Ashvin Vishwanath & Amir Yacoby

2. Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

Yonglong Xie, Jeong Min Park, Yuan Cao & Pablo Jarillo-Herrero

3. Research Center for Functional Materials, National Institute for Material Science, Tsukuba, Japan

Kenji Watanabe

4. International Center for Materials Nanoarchitectonics, National Institute for Material Science, Tsukuba, Japan

Takashi Taniguchi

### Contributions

Y.X., A.T.P., J.M.P., P.J.-H. and A.Y. conceived and designed the experiment. A.T.P. and Y.X. performed the scanning SET experiment and analysed the data with input from A.Y. J.M.P., Y.C. and P.J.-H. designed and provided the samples and contributed to the analysis of the results. D.E.P., E.K., P.L. and A.V. performed the theoretical analysis. K.W. and T.T. provided hBN crystals. All authors participated in discussions and in writing of the manuscript.

## Corresponding authors

Correspondence to [Yonglong Xie](#), [Pablo Jarillo-Herrero](#) or [Amir Yacoby](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended data Fig. 1 Additional CDW and SBCI states at higher magnetic field.](#)

**a–d**, Measurements of  $d\mu/dn$  ( $\times 10^{-11}$  mV cm $^{-2}$ ) in various density ranges between 6.5 and 12 T showing additional CDWs and SBCIs. **e–g**, Schematic Wannier diagrams corresponding to the states observed in **a–d** colored according to the classification used in Fig. [1b](#). Light blue and yellow lines denote the CDWs and SBCIs, respectively.

### [Extended Data Fig. 2 Energy gaps of additional FCI and SBCI states.](#)

**a–d**, Chemical potential steps  $\Delta\mu$  of the FCI (yellow and green circles) and SBCI (light and dark blue circles) states shown in Fig. [4](#).

### [Extended Data Fig. 3 Fits to \$\(t, s\)\$ for FCIs and SBCIs.](#)

**a–e**, Incompressible peak locations (blue circles) associated with FCI and SBCI states. Black lines mark the results of linear fits. The fitted slope of nearby Chern insulators were used to convert the parameters to  $(t, s)$ , the values of which are shown in the brackets with 95% confidence intervals.

### [Extended Data Fig. 4 Effect of twist angle inhomogeneity on FCIs.](#)

**a**, Local twist angle variation over a distance of  $1.6 \mu\text{m}$ . **b**, Compressibility measurements between  $\nu = 3$  and 4 measured along the same trajectory as in **a** at 9 T. The blue and black arrows indicate the densities near which the  $(2/3, 10/3)$  and  $(1/3, 11/3)$  states occur, respectively. **c**, Compressibility measurements taken at three locations indicated by the dotted lines in **a**. The white, blue, black and green dotted lines mark the expected evolution of the incompressible peaks associated with  $(1, 3)$ ,  $(2/3, 10/3)$ ,  $(1/3, 11/3)$  and  $(1, 13/4)$ .

## Extended Data Fig. 5 Hofstadter spectrum.

**a**, Calculated spectrum of the narrow bands of MATBG at finite magnetic field. The bands are colored according to their Chern number at zero magnetic field. **b**, Wannier plot corresponding to the spectrum in **a**.

# Supplementary information

## Supplementary Information

This file has two sections. The first describes the relation between the stability of FCIs and the quantum geometry of twisted BLG. The second section is entirely technical and gives the complete mathematical and numerical details of our models.

## Source data

### Source Data Fig. 1

### Source Data Fig. 2

### Source Data Fig. 3

### Source Data Fig. 4

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a

credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## [Reprints and Permissions](#)

## About this article

### Cite this article

Xie, Y., Pierce, A.T., Park, J.M. *et al.* Fractional Chern insulators in magic-angle twisted bilayer graphene. *Nature* **600**, 439–443 (2021).  
<https://doi.org/10.1038/s41586-021-04002-3>

- Received: 28 May 2021
- Accepted: 07 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04002-3>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04002-3>

- Article
- Open Access
- [Published: 15 December 2021](#)

# Tritiation of aryl thianthrenium salts with a molecular palladium catalyst

- [Da Zhao](#) ORCID: [orcid.org/0000-0002-7448-6969<sup>1</sup>](https://orcid.org/0000-0002-7448-6969),
- [Roland Petzold<sup>1</sup>](#),
- [Jiyao Yan](#) ORCID: [orcid.org/0000-0002-5122-2303<sup>1,2</sup>](https://orcid.org/0000-0002-5122-2303),
- [Dieter Muri](#) ORCID: [orcid.org/0000-0002-7617-9364<sup>3</sup>](https://orcid.org/0000-0002-7617-9364) &
- [Tobias Ritter](#) ORCID: [orcid.org/0000-0002-6957-450X<sup>1</sup>](https://orcid.org/0000-0002-6957-450X)

*Nature* volume 600, pages 444–449 (2021)

- 2540 Accesses
- 85 Altmetric
- [Metrics details](#)

## Subjects

- [Catalyst synthesis](#)
- [Homogeneous catalysis](#)
- [Synthetic chemistry methodology](#)

## Abstract

Tritium labelling is a critical tool for investigating the pharmacokinetic and pharmacodynamic properties of drugs, autoradiography, receptor binding and receptor occupancy studies<sup>1</sup>. Tritium gas is the preferred source of tritium for the preparation of labelled molecules because it is available in high isotopic purity<sup>2</sup>. The introduction of tritium labels from tritium gas is commonly achieved by heterogeneous transition-metal-catalysed tritiation of aryl (pseudo)halides. However, heterogeneous catalysts such as palladium supported on carbon operate through a reaction mechanism that also results in the reduction of other functional groups that are prominently featured in

pharmaceuticals<sup>3</sup>. Homogeneous palladium catalysts can react chemoselectively with aryl (pseudo)halides but have not been used for hydrogenolysis reactions because, after required oxidative addition, they cannot split dihydrogen<sup>4</sup>. Here we report a homogenous hydrogenolysis reaction with a well defined, molecular palladium catalyst. We show how the thianthrene leaving group—which can be introduced selectively into pharmaceuticals by late-stage C–H functionalization<sup>5</sup>—differs in its coordinating ability to relevant palladium(II) catalysts from conventional leaving groups to enable the previously unrealized catalysis with dihydrogen. This distinct reactivity combined with the chemoselectivity of a well defined molecular palladium catalyst enables the tritiation of small-molecule pharmaceuticals that contain functionality that may otherwise not be tolerated by heterogeneous catalysts. The tritiation reaction does not require an inert atmosphere or dry conditions and is therefore practical and robust to execute, and could have an immediate impact in the discovery and development of pharmaceuticals.

[Download PDF](#)

## Main

Tritium ( $^3\text{H}$ ) labelling allows the direct incorporation of a radioactive tag into pharmaceutical candidates without substantial changes in their chemical and physical properties and biological activity<sup>6</sup>. However, many reliable hydrogenation or hydrogenolysis reactions cannot be suitably used for tritium labelling owing to a lack of reagents, low molar activity, restricted functional-group tolerance or safety concerns. Tritiated water ( $^3\text{H}_2\text{O}$ ) is problematic owing to fast washout of the label from omnipresent water and safety concerns regarding the potential fast uptake of radioactive water by experimentalists. The preferred source of tritium labels is tritium gas ( $^3\text{H}_2$ ), which is available in high isotopic purity and practical to handle with commercially available manifolds on the small scale typically used for labelling<sup>2,7</sup>.

Hydrogenation with hydrogen gas is one of the most extensively studied reactions in chemistry, with numerous important applications ranging from biomass degradation<sup>8</sup> to hydrogenolysis of otherwise persistent halogenated pollutants<sup>9</sup>. Several well defined, homogeneous transition-metal catalysts based on rhodium<sup>10</sup>, iridium<sup>11</sup> and ruthenium<sup>12</sup> can split the strong hydrogen–hydrogen bond for countless productive hydrogenation reactions of unsaturated bonds. However, appropriate unsaturated bonds are often not present in pharmaceuticals or would be destroyed by hydrogenation, and the same hydrogenation catalysts are generally not useful for the hydrogenolysis of carbon–halide bonds because most transition-metal hydrides are inactive towards the oxidative addition of carbon–heteroatom bonds. In the presence of both dihydrogen and aryl(pseudo)halides, dihydrogen oxidative addition is

commonly faster, which results in metal hydrides in higher oxidation states that are not suitable for aryl (pseudo)halide oxidative addition<sup>13,14</sup>. Therefore, for the hydrogenolysis of carbon–heteroatom bonds<sup>15</sup>, chemists select heterogeneous catalysts, such as palladium supported on carbon, which can effectively reduce aryl(pseudo)halides through a mechanistically distinct pathway<sup>16,17</sup>. The reactivity of the active hydrogen chemisorbed on the catalyst surface results in low chemoselectivity, and the undesired reduction of other functional groups typically found in pharmaceuticals(Fig. 1a)<sup>3</sup>.

**Fig. 1: Palladium-catalysed hydrogenolysis with molecular hydrogen.**

---

 **figure1**

---

**a**, Heterogeneous Pd-catalysed hydrogenolysis with  $H_2$ . **b**, Homogeneous Pd-catalysed hydrogenolysis of aryl halide. L, neutral 2-electron ligand. **c**, Homogeneous Pd-catalysed hydrogenolysis of aryl thianthrenium salt. **d**, Chemo- and site-selective C–H tritiation via arylthianthrenium salt by homogeneous palladium catalysis. X, conventional (pseudo)halide.

Tritium for hydrogen exchange is a desirable way for tritium incorporation because prior functionalization is not required<sup>18</sup>. Several impressive hydrogen isotope

exchange reactions have been developed with transition-metal catalysts, maybe most notably those based on iridium<sup>19,20</sup> and nickel<sup>21</sup>; however, they require the presence of directing groups or heterocycles for efficient transformations. An aromatic hydrogen isotope exchange reaction that does not require directing groups was developed based on an iron catalyst, which provides successful tritiation complementary to the iridium-catalysed methods<sup>22</sup>. At present, the low-valent iron catalyst does not tolerate protic functional groups and requires an inert atmosphere, which complicates its routine, practical application in pharmaceutical development<sup>23,24</sup>. In addition, all hydrogen isotope exchange reactions typically afford several constitutional isomers with more than one label incorporated, which can be advantageous because high molar activity can be reached<sup>25</sup>, but may also render interpretation of imaging studies difficult when the labelled molecule is metabolized<sup>26</sup>. In that sense, a robust, practical, site-selective functionalization would be complementary to the hydrogen exchange reactions and could add significant value to the field<sup>27</sup>.

Metal-catalysed hydrogenolysis proceeds through metal-hydride intermediates<sup>28</sup> that can be formed by dihydrogen oxidative addition. Transition-metal complexes that react with aryl (pseudo)halides through oxidative addition do not engage with dihydrogen in a second oxidative reaction to form metal hydrides. Heterolytic cleavage of the hydrogen–hydrogen bond can also result in the formation of metal hydrides<sup>29</sup>. For example, electrophilic transition-metal complexes of iridium<sup>30</sup> and ruthenium<sup>31</sup> can substantially acidify dihydrogen through complexation owing to a strong  $\sigma$  donor–acceptor interaction<sup>29,32</sup>, which can result in the formation of metal hydrides through deprotonation of the metal  $\eta^2$ -dihydrogen complexes<sup>33</sup>. Only a few examples of palladium (Pd) dihydrogen complexes are known<sup>4</sup>, and no electrophilic Pd(II) catalysts—generated through oxidative addition of an aryl (pseudo)halide to low-valent palladium—have been shown to heterolyse dihydrogen<sup>34,35,36</sup>. Upon oxidative addition to a ligated Pd(0), no appropriate coordination site is available at square planar  $d^8$  Pd(II) tetracoordinate complexes for dihydrogen coordination because (pseudo)halides outcompete dihydrogen from coordination to Pd(II) (Fig. 1b). Aryl thianthrenium salts can be more reactive than aryl halides and aryl triflates in cross-coupling reactions<sup>5,37</sup> and photoredox catalysis<sup>38,39,40</sup>, and they are readily accessed by direct, late-stage C–H functionalization of complex small-molecule arenes in exquisite site selectivity. The fundamental novelty reported here entails that dihydrogen can be split by a palladium catalyst as part of a productive catalytic cycle that includes oxidative addition of aryl thianthrenium salts (Fig. 1c). The ability to engage structurally complex arenes and small-molecule pharmaceuticals at a late stage without directing groups, excellent functional group tolerance, exquisite site selectivity and high isotopic purity, in combination with operational simplicity, distinguish this protocol from heterogeneous hydrodehalogenation and hydrogen isotope exchange reactions for practical and robust tritiation (Fig. 1d).

Hydrogenolysis of aryl thianthrenium salts was performed with  $^2\text{H}_2$  as a more convenient and safer  $^3\text{H}_2$  surrogate. Palladium-catalysed hydrogenolysis of biphenyl-derived thianthrenium salt **1-TT** with  $^2\text{H}_2$  afforded the desired  $^2\text{H}$ -labelled product [ $^2\text{H}$ ]1 with more than 99% deuterium (D) incorporation (Fig. 2a). No synthetically useful reaction was observed with aryl bromides, aryl iodides and aryl triflates, nor with other arylsulfonium salts such as those derived from dibenzothiophene (DBT) and diphenyl sulfide (DPS), respectively. The lack of reactivity of the aryl halides could not be addressed through the use of other commonly used, highly active monodentate phosphine ligands (Fig. 2a, Supplementary Table 11). Likewise, the addition of (pseudo)halide anions to the productive reaction of arylthianthrenium salts poisons the reactivity, probably due to palladophilic halide outcompeting dihydrogen for binding; the starting material **1-TT** is recovered in these cases (Fig. 2b). Although sulfur-containing molecules are often responsible for catalyst poisoning owing to strong coordination<sup>41</sup>, we could not identify any coordination of thianthrene to cationic Pd(II), which is consistent with an open coordination site on the metal centre for interaction and coordination of  $\text{H}_2$  during catalysis. As measured by infrared spectroscopy, even triflate outcompetes thianthrene from coordination to Pd(II) (Supplementary Scheme 2, Supplementary Fig. 23). Cationic Pd(II) complexes can commonly be obtained from Pd(II) aryl halides by addition of silver(I) ( $\text{Ag}(\text{I})$ ) salts with non-coordinating anions during catalysis, as exemplified by Overman in Heck reactions that proceed through a mechanistically relevant cationic reaction pathway<sup>42</sup>. However, hydrogenolysis of aryl bromide in the presence of Ag(I) additives did not result in a productive reaction as observed with the arylthianthreniums, which may be due to the incompatibility of oxidizing Ag(I) and the presumed in situ-generated palladium hydride (Fig 2a).

**Fig. 2: Homogeneous palladium-catalysed reductive deuteration of aryl thianthrenium salt and aryl (pseudo)halides.**

---

 figure2

**a**, Hydrogenolysis of aryl (pseudo) halides. dppf, 1,1'-bis(diphenylphosphino)ferrocene. <sup>a</sup>For an additional evaluation of highly active monodentate phosphine ligands, see Supplementary Table 11. **b**, Effect of halide anions in hydrogenolysis of aryl thianthrenium salt.

(Tetrafluoro)thianthrenium (TFT and TT, respectively) salts of electron-deficient (**2**, **7**), electron-neutral (**4**, **14**, **24**) and electron-rich arenes (**3**, **8**, **11**), as well as arenes bearing *ortho* substituents (**15**, **17**, **22**) successfully participate in Pd-catalysed reductive deuteration (Fig. **3**). Functional-group compatibility is high, as evidenced by substrates containing unprotected hydroxyl groups (**4**), carbamates (**5**), amides (**10**, **20**, **25**), heterocycles (**6**, **13**, **18**) and acid-sensitive glycosidic linkages (**21**). In contrast to heterogeneous catalysis, the homogeneous catalyst system tolerates reducible functional groups such as aryl (pseudo)halides (**2**, **3**, **6**, **7**, **8**), nitriles (**3**), alkenes (**20**, **26**, **27**), esters (**14**, **15**), aldehydes (**23**), sulfones (**26**), nitro groups (**19**) and benzyl ethers (**23**). Primary amines must be protected. No  $^1\text{H}$ - $^2\text{H}$  scrambling was observed when the reaction was performed in the presence of water, and high levels of deuterium incorporation were detected throughout. When compared with heterogeneous deuterodehalogenation, where significant isotopic scrambling can occur<sup>[43](#)</sup>, the new catalytic platform with aryl thianthrenium salts provides a viable approach for the synthesis of deuterated and tritiated drug molecules (vide infra) with high isotopic purity in a practical setting that does not require the rigorous absence of water or oxygen.

---

**Fig. 3: Substrate scope for reductive deuteration of thianthrenium salts.**

---

 figure3

Deuterium-labelling of aryl thianthrenium salts. The general reaction conditions are as follows: (tetrafluoro)thianthrenium salt (0.20 mmol, 1.0 equiv.),  $K_3PO_4$  (1.0 equiv.),  $Pd(OAc)_2$  (1 mol%), dppf (1 mol%) or  $Pd[(PtBu_3)]_2$  (1 mol%), THF (0.1 M or 0.2 M),  $^2H_2$  (1 atm, about 5 equiv.), 23 °C, 12 h. <sup>a</sup>Yield of the isolated product. <sup>b</sup>Deuterium incorporation determined by  $^1H$  NMR. <sup>c</sup>H incorporation determined by mass analysis.

<sup>d</sup>5 mol% of catalyst. <sup>e</sup>2.5 mol% of catalyst. <sup>f</sup>48 h. <sup>g</sup>10 mol% of catalyst. <sup>h</sup>2 h. <sup>i</sup>Isolated as triflimide adduct. TFT, 2,3,7,8-tetrafluorothianthrene.

Given our design of electrophilic, cationic Pd(II) species for H<sub>2</sub> activation and the observed poisoning of the active catalyst by anions as weakly coordinating as triflate, successful reductive deuteration of substrates featuring coordinating groups such as tertiary amines (**6**, **20**), pyridines (**11**, **25**), pyrazoles (**18**), quinolines (**22**), anilines (**10**, **16**) and thiophenes (**17**) is non-trivial. We observed that the appropriate choice of counterion for the arylthianthrenium salt starting material can reduce its solubility in tetrahydrofuran (THF), with the reservoir of starting material in the solid phase, leaching material for conversion as the reaction progresses. The ability to tune solubility is another yet-unappreciated advantage of the arylthianthrenium salts when compared with arenes with conventional leaving groups that lack such a handle.

Translation of the hydrogenolysis reaction to tritium labelling was accomplished on micromole scales at a subatmospheric pressure of <sup>3</sup>H<sub>2</sub> gas to reduce the risk of tritium gas leakage and with a higher loading of catalyst to achieve faster reaction rates (Fig. **4**). No special care is required to exclude air or moisture during the radiosynthesis, and, in contrast to purification after hydrogen exchange reactions, the radiolabelled product can be readily separated from the starting material owing to the pronounced polarity difference due to the cationic nature of the thianthrenium salts. No isotope scrambling and straightforward purification result in high molar activity, which is often required for receptor binding and occupancy studies<sup>44</sup>. Higher molar activities can be achieved when more than one label is introduced, as shown for <sup>3</sup>H<sub>2</sub>-etofenprox ([<sup>3</sup>H<sub>2</sub>]**29**). In all cases, high, predictable positional selectivity enables single, well defined, labelled molecules that are generally not accessible with hydrogen isotope exchange reactions unless directing groups are used.

**Fig. 4: Substrate scope for reductive tritiation of thianthrenium salts.**

 figure4

Tritium labelling of aryl (tetrafluoro)thianthrenium salts. The general reaction conditions are as follows: (tetrafluoro)thianthrenium salt (3.0 μmol, 1.0 equiv.), K<sub>3</sub>PO<sub>4</sub>

(1.0 equiv.), Pd(OAc)<sub>2</sub> (10 mol%), dppf (10 mol%), THF (6 mM), <sup>3</sup>H<sub>2</sub> (0.3 atm, about 20 equiv.), 23 °C, 19 h. <sup>a</sup>68 h. <sup>b</sup>K<sub>3</sub>PO<sub>4</sub> (4 equiv.), <sup>3</sup>H<sub>2</sub> (0.7 atm, about 20 equiv.), 4 h.

A plausible reaction pathway that is consistent with all experimental data is outlined in Fig. 5. The generation of a catalytically active, monoligated Pd(0) catalyst from the only observed resting state Pd[(P*t*Bu<sub>3</sub>)<sub>2</sub>] during catalysis is consistent with the observed reaction order of 0.5 in Pd[(P*t*Bu<sub>3</sub>)<sub>2</sub>]. We determined a primary kinetic isotope effect (KIE) of  $k_{\text{H}}/k_{\text{D}} = 3.1$ , where  $k$  is the rate constant, by measuring the initial rates of independent reactions with H<sub>2</sub> and <sup>2</sup>H<sub>2</sub>, respectively, at the same pressure, and an equilibrium isotope effect of 1.1 from an intermolecular competition experiment between H<sub>2</sub> and <sup>2</sup>H<sub>2</sub> present at the same partial pressure, which is inconsistent with Ar-TT<sup>+</sup> oxidative addition preceding dihydrogen binding but consistent with reversible dihydrogen association before oxidative addition to Ar-TT<sup>+</sup>. The observed first order in Ar-TT<sup>+</sup> and positive order in H<sub>2</sub> are in agreement with this proposal. Although we cannot exclude dihydrogen oxidative addition from **A** based on our kinetic data, it would need to be reversible, as indicated by the competition KIE, and the Pd(0) intermediate **A** could be expected to react with Ar-TT<sup>+</sup> faster in concerted oxidative-addition or single-electron-transfer pathways than a putative off-cycle Pd(II) dihydride. Irrespective of the exact mechanism of irreversible—measured zero-order in TFT—oxidative addition, which may also differ depending on the ancillary ligand(s), intermediate **B** distinguishes the proposed pathway from those that would be accessible with other aryl(pseudo)halides. No H/D scrambling into the product was observed when hydrogenolysis was performed in the presence of H<sub>2</sub>O (Supplementary Figs. 4, 5), consistent with irreversible proton transfer from **B**. As the base is not fully dissolved in the liquid phase, its concentration does not appear in the rate law, but the primary KIE excludes turnover-limiting oxidative addition. Similarly, an inverse KIE or a primary KIE much smaller than 3.1 would be observed for turnover-limiting reductive elimination<sup>45</sup> from **C**; moreover, C-H reductive elimination is expected to be fast<sup>46,47</sup>. The data are consistent with turnover-limiting dihydrogen splitting to **C**, followed by fast reductive elimination to product. The addition of halides and pseudohalides or other soluble Lewis bases may prevent formation of **B** in sufficient quantities. Although aryldiazonium salts are typically not accessible through late-stage functionalization, a productive albeit less efficient (Supplementary Scheme 1), hydrogenolysis of biphenyldiazonium tetrafluoroborate under our reaction conditions is consistent with our mechanism hypothesis, as no coordinating anion is generated upon oxidative addition. Although some homogeneous transition-metal complexes can function as precursors for heterogeneous nanoparticles, we have excluded active heterogeneous catalysts as the major contributor to catalysis by means of observing the reaction in the presence and absence of mercury (Supplementary Fig. 22)<sup>48</sup>.

**Fig. 5: Plausible reaction pathway.**

 figure5

B, K<sub>3</sub>PO<sub>4</sub>. Counterion can also be PF<sub>6</sub> or NTf<sub>2</sub>.

In contrast to (pseudo)halides, the thianthrenyl group can be introduced selectively into complex small molecules. We show here how the lack of strong coordination to palladium and the intrinsic solubility properties enables thianthrenium chemistry to solve the challenge of homogeneous palladium-catalysed hydrogenolysis for the chemoselective synthesis of tritium-labelled small molecules.

## Methods

See Supplementary Information for further methods.

## Data availability

The data reported in this paper are available in the main text or the Supplementary Information.

## References

1. 1.

Isin, E. M., Elmore, C. S., Nilsson, G. N., Thompson, R. A. & Weidolf, L. Use of radiolabeled compounds in drug metabolism and pharmacokinetic studies. *Chem. Res. Toxicol.* **25**, 532–542 (2012).

2. 2.

Voges, R., Heys, J. R. & Moenius, T. in *Preparation of Compounds Labeled with Tritium and Carbon-14* 109–209 (John Wiley, 2009).

3. 3.

Nishimura S. in *Handbook of Heterogeneous Catalytic Hydrogenation for Organic Synthesis* 572–663 (Wiley-Interscience, 2001).

4. 4.

Connelly, S. J., Chanez, A. G., Kaminsky, W. & Heinekey, D. M. Characterization of a palladium dihydrogen complex. *Angew. Chem. Int. Ed.* **54**, 5915–5918 (2015).

5. 5.

Berger, F. et al. Site-selective and versatile aromatic C–H functionalization by thianthrenation. *Nature* **567**, 223–228 (2019).

6. 6.

Atzrodt, J., Derdau, V., Kerr, W. J. & Reid, M. Deuterium- and tritium-labelled compounds: applications in the life sciences. *Angew. Chem. Int. Ed.* **57**, 1758–1784 (2018).

7. 7.

Saljoughian, M. Synthetic tritium labeling: reagents and methodologies. *Synthesis* **13**, 1781–1801 (2002).

8. 8.

Yang, Y. & McCarty, P. L. Biomass, oleate, and other possible substrates for chloroethene reductive dehalogenation. *Bioremediat. J.* **4**, 125–133 (2000).

9. 9.

Nakano, M. M. & Zuber, P. in *Strict and Facultative Anaerobes Medical and Environmental Aspects* 303–317 (CRC Press, 2004).

10. 10.

Shapley, J. R., Schrock, R. R. & Osborn, J. A. Preparation and catalytic properties of some cationic iridium(III) and rhodium(III) dihydrido complexes. *J. Am. Chem. Soc.* **91**, 2816–2817 (1969).

11. 11.

Vaska, L. & DiLuzio, J. W. Activation of hydrogen by a transition metal complex at normal conditions leading to a stable molecular dihydride. *J. Am. Chem. Soc.* **84**, 679–680 (1962).

12. 12.

Chinn, M. S. & Heinekey, D. M. Synthesis and properties of a series of ruthenium dihydrogen complexes. *J. Am. Chem. Soc.* **109**, 5865–5867 (1987).

13. 13.

Fan, L., Parkin, S. & Ozerov, O. V. Halobenzenes and Ir(I): kinetic C–H oxidative addition and thermodynamic C–Hal oxidative addition. *J. Am. Chem. Soc.* **127**, 16772–16773 (2005).

14. 14.

Douglas, T. M., Chaplin, A. B. & Weller, A. S. Dihydrogen loss from a 14-electron rhodium(III) bis-phosphine dihydride to give a rhodium(I) complex that undergoes oxidative addition with aryl chlorides. *Organometallics* **27**, 2918–2921 (2008).

15. 15.

Alonso, F. Beletskaya, I. P. & Yus, M. Metal-mediated reductive hydrodehalogenation of organic halides. *Chem. Rev.* **102**, 4009–4091 (2002).

16. 16.

van Santen, R. A. in *Modern Heterogeneous Catalysis: An Introduction* 293–344 (John Wiley, 2017).

17. 17.

Sisak, A. & Simon O. B. in *The Handbook of Homogeneous Hydrogenation* (eds de Vries J. G. & Elsevier, C. J.) 513–546 (Wiley-VCH, 2007).

18. 18.

Atzrodt, J., Derdau, V., Kerr, W. J. & Reid, M. C–H functionalisation for hydrogen isotope exchange. *Angew. Chem. Int. Ed.* **57**, 3022–3047 (2018).

19. 19.

Heys, R. Investigation of  $[\text{IrH}_2(\text{Me}_2\text{CO})_2(\text{PPh}_3)_2]\text{BF}_4$  as a catalyst of hydrogen isotope exchange of substrates in solution. *J. Chem. Soc. Chem. Commun.* 680–681 (1992).

20. 20.

Ellames, G. J., Gibson, J. S., Herbert, J. M.; Kerr, W. J. & McNeill, A. H. Deuterium exchange mediated by an iridium–phosphine complex formed in situ. *Tetrahedron Lett.* **42**, 6413–6416 (2001).

21. 21.

Zarate, C., Yang, H., Bezdek, M. J., Hesk, D. & Chirik, P. J. Ni(I)–X complexes bearing a bulky  $\alpha$ -diimine ligand: synthesis, structure, and superior catalytic performance in the hydrogen isotope exchange in pharmaceuticals. *J. Am. Chem. Soc.* **141**, 5034–5044 (2019).

22. 22.

Yang, H. et al. Site-selective nickel-catalyzed hydrogen isotope exchange in N-heterocycles and its application to the tritiation of pharmaceuticals. *ACS Catal.* **8**, 10210–10218 (2018).

23. 23.

Yang, H. & Hesk, D. Base metal-catalyzed hydrogen isotope exchange. *J. Labelled Compd Radiopharm.* **63**, 296–307 (2020).

24. 24.

Yu, R. P., Hesk, D., Rivera, N., Pelczer, I. & Chirik, P. J. Iron-catalysed tritiation of pharmaceuticals. *Nature* **529**, 195–199 (2016).

25. 25.

Yang, H., Dormer, P. G., Rivera, N. R. & Hoover, A. J. Palladium (II)-mediated C–H tritiation of complex pharmaceuticals. *Angew. Chem. Int. Ed.* **57**, 1883–1887 (2018).

26. 26.

Lockley, W. J. S., McEwen, A. & Cooke, R. Tritium: a coming of age for drug discovery and development ADME studies. *J. Labelled Compd Radiopharm.* **55**, 235–257 (2012).

27. 27.

Koniarczyk, J. L., Hesk, D., Overgard, A., Davies, I. W. & McNally, A. A general strategy for site-selective incorporation of deuterium and tritium into pyridines, diazines, and pharmaceuticals. *J. Am. Chem. Soc.* **140**, 1990–1993 (2018).

28. 28.

Sean, M. G. & Gemma, G. The multifarious world of transition metal hydrides. *Chem. Soc. Rev.* **32**, 383–392 (2003).

29. 29.

Kubas, G. J. Fundamentals of H<sub>2</sub> binding and reactivity on transition metals underlying hydrogenase function and H<sub>2</sub> production and storage. *Chem. Rev.* **107**, 4152–4205 (2007).

30. 30.

Crabtree, R. H., Lavin, M. & Bonneviot, L. J. Some molecular hydrogen complexes of iridium *J. Am. Chem. Soc.* **108**, 4032–4037 (1986).

31. 31.

Gunanathan, C. & Milstein, D. Bond activation and catalysis by ruthenium pincer complexes. *Chem. Rev.* **114**, 12024–12087 (2014).

32. 32.

Crabtree, R. H. Dihydrogen complexation. *Chem. Rev.* **116**, 8750–8769 (2016).

33. 33.

Heinekey, D. M. & Oldham, W. J. Coordination chemistry of dihydrogen *Chem. Rev.* **93**, 913–926 (1993).

34. 34.

Fulmer, G. R., Muller, R. P., Kemp, R. A. & Goldberg, K. I. Hydrogenolysis of palladium(II) hydroxide and methoxide pincer complexes. *J. Am. Chem. Soc.* **131**, 1346–1347 (2009).

35. 35.

Fulmer, G. R., Herndon, A. N., Kaminsky, W., Kemp, R. A. & Goldberg, K. I. Hydrogenolysis of palladium(II) hydroxide, phenoxide, and alkoxide complexes. *J. Am. Chem. Soc.* **133**, 17713–17726 (2011).

36. 36.

Leñero, K. Q. et al. A heterolytic activation of dihydrogen by platinum and palladium complexes. *Dalton Trans.* **42**, 6495–6512 (2013).

37. 37.

Engl, P. S. et al. C–N cross-couplings for site-selective late-stage diversification via aryl sulfonium salts. *J. Am. Chem. Soc.* **141**, 13346–13351 (2019).

38. 38.

Li, J. et al. Photoredox catalysis with aryl sulfonium salts enables site-selective late-stage fluorination. *Nat. Chem.* **12**, 56–62 (2020).

39. 39.

Ye, F. et al. Aryl sulfonium salts for site-selective late-stage trifluoromethylation. *Angew. Chem. Int. Ed.* **58**, 14615–14619 (2019).

40. 40.

Sang, R. et al. Site-selective C–H oxygenation via aryl sulfonium salts. *Angew. Chem. Int. Ed.* **58**, 16161–16166 (2019).

41. 41.

Wise, H. Mechanisms of catalyst poisoning by sulfur species. *Stud. Surf. Sci. Catal.* **68**, 497–504 (1991).

42. 42.

Abelman, M. M., Oh, T. & Overman, L. E. Intramolecular alkene arylations for rapid assembly of polycyclic systems containing quaternary centers. A new synthesis of spirooxindoles and other fused and bridged ring systems. *J. Org. Chem.* **52**, 4130–4133 (1987).

43. 43.

Lockley, W. J. S., Venanzi, N. A. E. & Crane, G. J. Studies of hydrogen isotope scrambling during the dehalogenation of aromatic chloro-compounds with deuterium gas over palladium catalysts. *J. Labelled Compd Radiopharm.* **63**, 531–552 (2020).

44. 44.

Hein, P. et al. in *Practical Methods in Cardiovascular Research* (eds Dhein, S. et al.) 723–783 (Springer, 2005).

45. 45.

Churchill, D. G., Janak, K. E., Wittenberg, J. S. & Parkin, G. Normal and inverse primary kinetic deuterium isotope effects for C–H bond reductive elimination and oxidative addition reactions of molybdenocene and tungstenocene complexes: evidence for benzene sigma-complex intermediates. *J. Am. Chem. Soc.* **125**, 1403–1420 (2003).

46. 46.

Low, J. J. & Gaddard, W. A. Reductive coupling of hydrogen–hydrogen, hydrogen–carbon, and carbon–carbon bonds from palladium complexes. *J. Am. Chem. Soc.* **106**, 8321–8322 (1984).

47. 47.

Low, J. J. & Gaddard, W. A. Theoretical studies of oxidative addition and reductive elimination. 3. Carbon–hydrogen and carbon–carbon reductive coupling from palladium and platinum bis(phosphine) complexes. *J. Am. Chem. Soc.* **108**, 6115–6128 (1986).

48. 48.

Widegren, J. A. & Finke, R. G. A review of the problem of distinguishing true homogeneous catalysis from soluble or other metal-particle heterogeneous catalysis under reducing conditions. *J. Mol. Catal. A* **198**, 317–341 (2003).

## Acknowledgements

We thank S. Marcus, D. Margold, F. Köhler, N. Haupt and D. Kampen for mass spectrometry analysis; M. Kochius, M. Leutzsch, C. Fares, S. Tobegen and C. Wirtz for NMR spectroscopy analysis; F. Kaegi (F. Hoffmann-La Roche Ltd) and RC Tritec for tritium-labelling experiments; and P. Münstermann for high-performance liquid chromatography analysis. We thank Q. Cheng, B. Lansbergen, F. Juliá and L. Zhang for helpful discussions.

## Funding

Open access funding provided by Max Planck Society.

## Author information

### Affiliations

1. Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany

Da Zhao, Roland Petzold, Jiyao Yan & Tobias Ritter

2. Institute of Organic Chemistry, RWTH Aachen University, Aachen, Germany

Jiyao Yan

3. Pre-clinical CMC, Roche Pharma Research and Early Development, Roche Innovation Center Basel, Basel, Switzerland

Dieter Muri

## Contributions

D.Z. developed the reaction chemistry and investigated the mechanism. D.Z., R.P. and J.Y. optimized and explored the substrate scope for reductive deuteration. D.M. optimized and performed the reductive tritiation reaction. D.Z., D.M. and T.R. wrote the manuscript. T.R. directed the project.

## Corresponding author

Correspondence to [Tobias Ritter](#).

## Ethics declarations

### Competing interests

T.R. may benefit from royalty payments related to sales from thianthrene-based compounds.

## Additional information

**Peer review information** *Nature* thanks Viktoria Däschlein-Gessner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

### [Supplementary Information](#)

This file contains detailed experimental procedures and spectroscopic data, and includes 11 supplementary tables and 33 supplementary figures.

### [Peer Review File](#)

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and

reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## Reprints and Permissions

## About this article

### Cite this article

Zhao, D., Petzold, R., Yan, J. *et al.* Tritiation of aryl thianthrenium salts with a molecular palladium catalyst. *Nature* **600**, 444–449 (2021).  
<https://doi.org/10.1038/s41586-021-04007-y>

- Received: 23 April 2021
- Accepted: 06 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04007-y>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 15 December 2021](#)

# A large West Antarctic Ice Sheet explains early Neogene sea-level amplitude

- [J. W. Marschalek](#) ORCID: [orcid.org/0000-0003-2057-4012](https://orcid.org/0000-0003-2057-4012)<sup>1</sup>,
- [L. Zurli](#) ORCID: [orcid.org/0000-0002-7669-7305](https://orcid.org/0000-0002-7669-7305)<sup>2</sup>,
- [F. Talarico](#)<sup>2 na1</sup>,
- [T. van de Flierdt](#) ORCID: [orcid.org/0000-0001-7176-9755](https://orcid.org/0000-0001-7176-9755)<sup>1</sup>,
- [P. Vermeesch](#)<sup>3</sup>,
- [A. Carter](#) ORCID: [orcid.org/0000-0002-0090-5868](https://orcid.org/0000-0002-0090-5868)<sup>4</sup>,
- [F. Beny](#) ORCID: [orcid.org/0000-0002-2322-4299](https://orcid.org/0000-0002-2322-4299)<sup>5</sup>,
- [V. Bout-Roumazeilles](#)<sup>5</sup>,
- [F. Sangiorgi](#) ORCID: [orcid.org/0000-0003-4233-6154](https://orcid.org/0000-0003-4233-6154)<sup>6</sup>,
- [S. R. Hemming](#) ORCID: [orcid.org/0000-0001-8117-2303](https://orcid.org/0000-0001-8117-2303)<sup>7</sup>,
- [L. F. Pérez](#) ORCID: [orcid.org/0000-0002-6229-4564](https://orcid.org/0000-0002-6229-4564)<sup>8,9</sup>,
- [F. Colleoni](#) ORCID: [orcid.org/0000-0003-4582-812X](https://orcid.org/0000-0003-4582-812X)<sup>10</sup>,
- [J. G. Prebble](#)<sup>11</sup>,
- [T. E. van Peer](#) ORCID: [orcid.org/0000-0003-3516-4198](https://orcid.org/0000-0003-3516-4198)<sup>3,12</sup>,
- [M. Perotti](#) ORCID: [orcid.org/0000-0003-4713-7867](https://orcid.org/0000-0003-4713-7867)<sup>2</sup>,
- [A. E. Shevenell](#) ORCID: [orcid.org/0000-0002-6457-6530](https://orcid.org/0000-0002-6457-6530)<sup>13</sup>,
- [I. Browne](#) ORCID: [orcid.org/0000-0002-1569-8208](https://orcid.org/0000-0002-1569-8208)<sup>13</sup>,
- [D. K. Kulhanek](#) ORCID: [orcid.org/0000-0002-2156-6383](https://orcid.org/0000-0002-2156-6383)<sup>14,15</sup>,
- [R. Levy](#) ORCID: [orcid.org/0000-0002-8783-0167](https://orcid.org/0000-0002-8783-0167)<sup>11,15</sup>,
- [D. Harwood](#) ORCID: [orcid.org/0000-0002-9449-2127](https://orcid.org/0000-0002-9449-2127)<sup>17</sup>,
- [N. B. Sullivan](#)<sup>18</sup>,
- [S. R. Meyers](#)<sup>18</sup>,
- [E. M. Griffith](#) ORCID: [orcid.org/0000-0002-7919-4551](https://orcid.org/0000-0002-7919-4551)<sup>19</sup>,

- [C.-D. Hillenbrand](#) [ORCID: orcid.org/0000-0003-0240-7317<sup>8</sup>](#),
- [E. Gasson<sup>20</sup>](#),
- [M. J. Siegert](#) [ORCID: orcid.org/0000-0002-0090-4806<sup>1,21</sup>](#),
- [B. Keisling](#) [ORCID: orcid.org/0000-0002-2182-2025<sup>7</sup>](#),
- [K. J. Licht<sup>22</sup>](#),
- [G. Kuhn](#) [ORCID: orcid.org/0000-0001-6069-7485<sup>23</sup>](#),
- [J. P. Dodd<sup>24</sup>](#),
- [C. Boshuis<sup>6</sup>](#),
- [L. De Santis](#) [ORCID: orcid.org/0000-0002-7752-7754<sup>10</sup>](#),
- [R. M. McKay](#) [ORCID: orcid.org/0000-0002-5602-6985<sup>16</sup>](#) &
- [IODP Expedition 374](#)

*Nature* volume **600**, pages 450–455 (2021)

- 978 Accesses
- 148 Altmetric
- [Metrics details](#)

## Subjects

- [Cryospheric science](#)
- [Geochemistry](#)
- [Palaeoclimate](#)

## Abstract

Early to Middle Miocene sea-level oscillations of approximately 40–60 m estimated from far-field records<sup>1,2,3</sup> are interpreted to reflect the loss of virtually all East Antarctic ice during peak warmth<sup>2</sup>. This contrasts with ice-sheet model experiments suggesting most terrestrial ice in East Antarctica was retained even during the warmest intervals of the Middle Miocene<sup>4,5</sup>. Data and model outputs can be reconciled if a large West Antarctic Ice Sheet (WAIS) existed and expanded across most of the outer continental shelf during the Early Miocene, accounting for maximum ice-sheet

volumes. Here we provide the earliest geological evidence proving large WAIS expansions occurred during the Early Miocene (~17.72–17.40 Ma). Geochemical and petrographic data show glacimarine sediments recovered at International Ocean Discovery Program (IODP) Site U1521 in the central Ross Sea derive from West Antarctica, requiring the presence of a WAIS covering most of the Ross Sea continental shelf. Seismic, lithological and palynological data reveal the intermittent proximity of grounded ice to Site U1521. The erosion rate calculated from this sediment package greatly exceeds the long-term mean, implying rapid erosion of West Antarctica. This interval therefore captures a key step in the genesis of a marine-based WAIS and a tipping point in Antarctic ice-sheet evolution.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

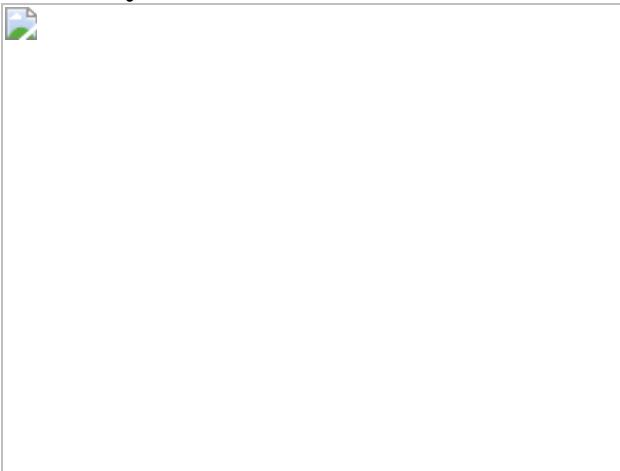
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

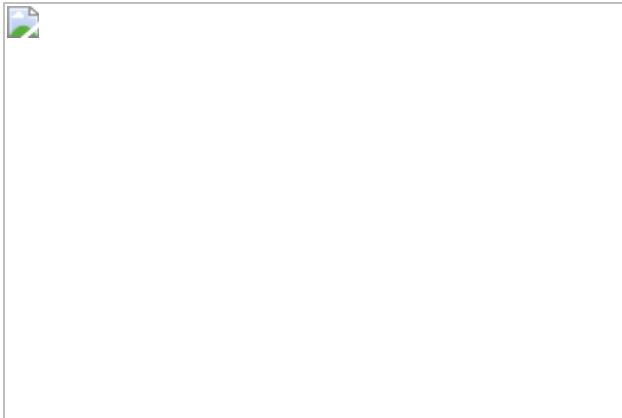
**Fig. 1: Site U1521 location and surrounding geology.**



**Fig. 2: Selected provenance proxies from IODP Site U1521 compared with Early Miocene climate records.**



**Fig. 3: Site U1521 detrital zircon U–Pb age distributions.**



## Data availability

The data sets generated as part of this study are available in the British Geological Survey National Geoscience Data Centre. Data sets include Nd and Sr isotope data (<https://doi.org/10.5285/3a646c8a-8422-4079-a928-a159532439eb>), zircon U-Pb dates (<https://doi.org/10.5285/cfadf931-0804-484c-a9d0-96254239c421>), clast counts (<https://doi.org/10.5285/b043471f-22e5-40e4-b274-1c875316d725>), clay mineralogy data (<https://doi.org/10.5285/b3cb3574-49b0-44c8-a934-3da88ca4ef93>), hornblende  $^{40}\text{Ar}/^{39}\text{Ar}$  dates (<https://doi.org/10.5285/926cad28-669f-4703-8a5b-5e7e843a4ee1>) and palynological counts (<https://doi.org/10.5285/adea0809-5fe5-4fb5-9f3e-9d774534d26d>). Source data are provided with this paper.

## References

1. 1.

Kominz, M. A. et al. Miocene relative sea level on the New Jersey shallow continental shelf and coastal plain derived from one-dimensional backstripping: A case for both eustasy and epeirogeny. *Geosphere* **12**, 1437–1456 (2016).

2. 2.

Miller, K. G. et al. Cenozoic sea-level and cryospheric evolution from deep-sea geochemical and continental margin records. *Sci. Adv.* **6**, eaaz1346 (2020).

3. 3.

Pekar, S. F. & DeConto, R. M. High-resolution ice-volume estimates for the early Miocene: Evidence for a dynamic ice sheet in Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 101–109 (2006).

4. 4.

Gasson, E., DeConto, R. M., Pollard, D. & Levy, R. H. Dynamic Antarctic ice sheet during the early to mid-Miocene. *Proc. Natl Acad. Sci. USA* **113**, 3459–3464 (2016).

5. 5.

Paxman, G. J., Gasson, E. G., Jamieson, S. S., Bentley, M. J., & Ferraccioli, F. Long-term increase in antarctic ice sheet vulnerability driven by bed topography evolution. *Geophys. Res. Lett.* **47**, e2020GL090003 (2020).

6. 6.

Masson-Delmotte, V. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 383–464 (IPCC, Cambridge Univ. Press, 2013).

7. 7.

Kennicutt, M. C. et al. A roadmap for Antarctic and Southern Ocean science for the next two decades and beyond. *Antarct. Sci.* **27**, 3–18 (2014).

8. 8.

Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoceanography. *J.*

*Geophys. Res.* **82**, 3843–3860 (1977).

9. 9.

Barrett, P. J. Characteristics of pebbles from Cenozoic marine glacial sediments in the Ross Sea (DSDP Sites 270–274) and the South Indian Ocean (Site 268). *Initial Rep. Deep Sea Drill. Proj.* **28**, 769–784 (1975).

10. 10.

Passchier, S. & Krissek, L. A. Oligocene–Miocene Antarctic continental weathering record and paleoclimatic implications, Cape Roberts drilling project, Ross Sea, Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 30–40 (2008).

11. 11.

Levy, R. et al. Antarctic ice sheet sensitivity to atmospheric CO<sub>2</sub> variations in the early to mid-Miocene. *Proc. Natl Acad. Sci. USA* **113**, 3453–3458 (2016).

12. 12.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).

13. 13.

Kennett, J. P. & Barker, P. F. Latest Cretaceous to Cenozoic climate and oceanographic developments in the Weddell Sea, Antarctica: an ocean-drilling perspective. *Proc. Ocean Drill. Program, Sci. Results* **113**, 937–960 (1990).

14. 14.

Hauptvogel, D. W. & Passchier, S. Early–Middle Miocene (17–14 Ma) Antarctic ice dynamics reconstructed from the heavy mineral provenance in the AND-2A drill core, Ross Sea, Antarctica. *Global Planet. Change* **82**, 38–50 (2012).

15. 15.

Levy, R. H. et al. Antarctic ice-sheet sensitivity to obliquity forcing enhanced through ocean connections. *Nat. Geosci.* **12**, 132–137 (2019).

16. 16.

Colleoni, F. et al. Past continental shelf evolution increased Antarctic ice sheet sensitivity to climatic conditions. *Sci. Rep.* **8**, 1–12 (2018).

17. 17.

Wilson, D. S. et al. Antarctic topography at the Eocene-Oligocene boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **335-336**, 24–34 (2012).

18. 18.

Paxman, G. J. et al. Reconstructions of Antarctic topography since the Eocene–Oligocene boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **535**, 109346 (2019).

19. 19.

Gasson, E. G. & Keisling, B. A. The Antarctic Ice Sheet: A Paleoclimate modelling perspective. *Oceanography (Wash. D.C.)* **33**, 90–100 (2020).

20. 20.

Anderson, J. B. & Bartek, L. R. in *The Antarctic Paleoenvironment: A Perspective on Global Change. Part One* Vol. 56 (eds Kennett, J.P & Warkne, D. A.) 231–264 (AGU, 1992).

21. 21.

De Santis, L., Anderson, J. B., Brancolini, G. & Zayatz, I. in *Geology and Seismic Stratigraphy of the Antarctic Margin* Vol. 68 (eds Cooper, A. K., Barker, P. F. & Brancolini, G.) 235–260 (AGU, 1995).

22. 22.

Gohl, K. et al. Seismic stratigraphic record of the Amundsen Sea Embayment shelf from pre-glacial to recent times: Evidence for a dynamic West Antarctic Ice Sheet. *Mar. Geol.* **344**, 115–131 (2013).

23. 23.

Pérez, L. F. et al. Early-middle Miocene ice sheet dynamics in the Ross Sea embayment: results from integrated core-log-seismic interpretation. *Geol. Soc. Am. Bull.* <https://doi.org/10.1130/B35814.1> (2021).

24. 24.

Bart, P. J. Were West Antarctic ice sheet grounding events in the Ross Sea a consequence of East Antarctic ice sheet expansion during the middle Miocene? *Earth Planet. Sci. Lett.* **216**, 93–107 (2003).

25. 25.

Chow, J. M. & Bart, P. J. West Antarctic Ice Sheet grounding events on the Ross Sea outer continental shelf during the middle Miocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **198**, 169–186 (2003).

26. 26.

McKay, R., De Santis, L. & Kulhanek, D. K. and the Expedition 374 Science Party. Ross Sea West Antarctic Ice Sheet History in *Proc. Int. Ocean Discovery Program* (IODP, 2019).

27. 27.

Licht, K. J. & Hemming, S. R. Analysis of Antarctic glaciogenic sediment provenance through geochemical and petrologic applications. *Quat. Sci. Rev.* **164**, 1–24 (2017).

28. 28.

Farmer, G. L., Licht, K., Swope, R. J. & Andrews, J. Isotopic constraints on the provenance of fine-grained sediment in LGM tills from the Ross Embayment, Antarctica. *Earth Planet. Sci. Lett.* **249**, 90–107 (2006).

29. 29.

van Wyck de Vries, M., Bingham, R. G. & Hein, A. S. in *Exploration of Subsurface Antarctica: Uncovering Past Changes and Modern Processes* (eds Siegert, M. J., Jamieson, S. S. R. & White, D. A.) <https://doi.org/10.1144/SP461.7> (Geological Society, 2017).

30. 30.

Farmer, G. L. & Licht, K. J. Generation and fate of glacial sediments in the central Transantarctic Mountains based on radiogenic isotopes and implications for reconstructing past ice dynamics. *Quat. Sci. Rev.* **150**, 98–109 (2016).

31. 31.

Goodge, J. W. Geological and tectonic evolution of the Transantarctic Mountains, from ancient craton to recent enigma. *Gondwana Res.* **80**, 50–122 (2020).

32. 32.

Licht, K. J. & Palmer, E. F. Erosion and transport by Byrd Glacier, Antarctica during the last glacial maximum. *Quat. Sci. Rev.* **62**, 32–48 (2013).

33. 33.

Licht, K. J., Hennessy, A. J. & Welke, B. M. The U-Pb detrital zircon signature of West Antarctic ice stream tills in the Ross embayment, with implications for Last Glacial Maximum ice flow reconstructions. *Antarct. Sci.* **26**, 687–697 (2014).

34. 34.

Bader, N. A., Licht, K. J., Kaplan, M. R., Kassab, C. & Winckler, G. East Antarctic ice sheet stability recorded in a high-elevation ice-cored moraine. *Quat. Sci. Rev.* **159**, 88–102 (2017).

35. 35.

Kyle, R. A. & Schopf, J. M. in *Antarctic Geoscience* (ed. Craddock, C.) 649–659 (Univ. Wisconsin Press, 1982).

36. 36.

Perotti, M., Andreucci, B., Talarico, F., Zattin, M. & Langone, A. Multianalytical provenance analysis of Eastern Ross Sea LGM till sediments (Antarctica): Petrography, geochronology, and thermochronology detrital data. *Geochem. Geophys. Geosyst.* **18**, 2275–2304 (2017).

37. 37.

Jordan, T. A., Riley, T. R. & Siddoway, C. S. The geological history and evolution of West Antarctica. *Nat. Rev. Earth Environ.* **1**, 117–133 (2020).

38. 38.

Balshaw-Biddle, K. M. *Antarctic Glacial Chronology Reflected in the Oligocene through Pliocene Sedimentary Section in the Ross Sea* PhD Thesis, Rice University (1981).

39. 39.

Westerhold, T. et al. An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science* **369**, 1383–1387 (2020).

40. 40.

Koppes, M. et al. Observed latitudinal variations in erosion as a function of glacier dynamics. *Nature* **526**, 100–103 (2015).

41. 41.

Alley, R. B., Cuffey, K. M. & Zoet, L. K. Glacial erosion: status and outlook. *Ann. Glaciol.* **60**, 1–13 (2019).

42. 42.

Cox, S. C., Smith Lyttle, B. and the GeoMAP team. SCAR GeoMAP dataset. GNS Science, Lower Hutt, New Zealand. Release v.201907  
<https://doi.org/10.21420/7SH7-6K05> (2019).

43. 43.

Morlighem, M. *MEaSUREs BedMachine Antarctica, Version 1* (Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2019; accessed 10 June 2021);  
<https://doi.org/10.5067/C2GFER6PTOS4>.

44. 44.

Morlighem, M. et al. Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet. *Nat. Geosci.* **13**, 132–137 (2020).

45. 45.

Mouginot, J., Scheuchl, B. & Rignot, E. *MEaSUREs Antarctic Boundaries for IPY 2007–2009 from Satellite Radar, Version 2* (Boulder, Colorado USA. NASA National Snow and Ice Data Center

Distributed Active Archive Center, 2017; accessed 12 June 2020);  
<https://doi.org/10.5067/AXE4121732AD>.

46. 46.

Rignot, E., Jacobs, S. S., Mouginot, J. & Scheuchl, B. Ice-shelf melting around. *Antarct. Sci.* **341**, 266–270 (2013).

47. 47.

Tinto, K. J. et al. Ross Ice Shelf response to climate driven by the tectonic imprint on seafloor bathymetry. *Nat. Geosci.* **12**, 441–449 (2019).

48. 48.

Vermeesch, P. Statistical models for point-counting data. *Earth Planet. Sci. Lett.* **501**, 112–118 (2018).

49. 49.

Ogg, J. *Geomagnetic Polarity Time Scale. In Geologic Time Scale 2020* (eds Gradstein, F. M. et al.) 159–192 (Elsevier, 2020).

50. 50.

Rae, J. W. et al. Atmospheric CO<sub>2</sub> over the past 66 million years from marine archives. *Annu. Rev. Earth Planet. Sci.* **49**, 599–631 (2021).

51. 51.

Vermeesch, P. Multi-sample comparison of detrital age distributions. *Chem. Geol.* **341**, 140–146 (2013).

52. 52.

Goldstein, S. L. & Hemming, S. R. in *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) 453–489 (Pergamon, 2003).

53. 53.

Garçon, M., Chauvel, C., France-Lanord, C., Huyghe, P. & Lavé, J. Continental sedimentary processes decouple Nd and Hf isotopes. *Geochim. Cosmochim. Acta* **121**, 177–195 (2013).

54. 54.

Gutjahr, M. et al. Reliable extraction of a deepwater trace metal isotope signal from Fe–Mn oxyhydroxide coatings of marine sediments. *Chem. Geol.* **242**, 351–370 (2007).

55. 55.

Simões Pereira, P. et al. Geochemical fingerprints of glacially eroded bedrock from West Antarctica: D–erital thermochronology, radiogenic isotope systematics and trace element geochemistry in Late Holocene glacial-marine sediments. *Earth Sci. Rev.* **182**, 204–232 (2018).

56. 56.

Tanaka, T. et al. JNdi-1: a neodymium isotopic reference in consistency with LaJolla neodymium. *Chem. Geol.* **168**, 279–281 (2000).

57. 57.

Weis, D. et al. High-precision isotopic characterization of USGS reference materials by TIMS and MC-ICP-MS. *Geochem. Geophys. Geosyst.* **7**, Q08006 (2006).

58. 58.

Jacobsen, S. B. & Wasserburg, G. J. Sm-Nd isotopic evolution of chondrites. *Earth Planet. Sci. Lett.* **50**, 139–155 (1980).

59. 59.

Sláma, J. et al. Plešovice zircon—a new natural reference material for UPb and Hf isotopic microanalysis. *Chem. Geol.* **249**, 1–35 (2008).

60. 60.

Pearce, N. J. et al. A compilation of new and published major and trace element data for NIST SRM 610 and NIST SRM 612 glass reference materials. *Geostand. Newslet.* **21**, 115–144 (1997).

61. 61.

Griffin, W. L. in *Laser Ablation ICP-MS in the Earth Sciences: Current Practices and Outstanding Issues* (ed. Sylvester, P.) 308–311 (Mineralogical Association of Canada, 2008).

62. 62.

Vermeesch, P. How many grains are needed for a provenance study? *Earth Planet. Sci. Lett.* **224**, 441–451 (2004).

63. 63.

Jackson, S. E., Pearson, N. J., Griffin, W. L. & Belousova, E. A. The application of laser ablation-inductively coupled plasma-mass spectrometry to in situ U–Pb zircon geochronology. *Chem. Geol.* **211**, 47–69 (2004).

64. 64.

Vermeesch, P. & Isoplot, R. A free and open toolbox for geochronology. *Geoscience Frontiers* **9**, 1479–1493 (2018).

65. 65.

Talarico, F. & Sandroni, S. Petrography, mineral chemistry and provenance of basement clasts in the CRP-1 drillcore (Victoria Land Basin, Antarctica). *Terra Antarct.* **5**, 601–610 (1998).

66. 66.

Talarico, F. & Sandroni, S. Provenance signature of the Antarctic Ice Sheets in the Ross Embayment during the Late Miocene to Early Pliocene: the ANDRILL AND-1B core record. *Global Planet. Change* **69**, 103–123 (2009).

67. 67.

Talarico, F., Sandroni, S., Fielding, C. R. & Atkins, C. Variability, petrography and provenance of basement clasts from CRP-2/2A drillcore (Victoria Land Basin, Ross Sea, Antarctica). *Terra Antarct.* **7**, 529–544 (2000).

68. 68.

Sandoni, S. & Talarico, F. M. Petrography and provenance of basement clasts and clast variability in CRP-3 drillcore (Victoria Land Basin, Antarctica). *Terra Antarct.* **8**, 449–467 (2001).

69. 69.

Wood, G. D., Gabriel, A. M. & Lawson, J. C. in *Palynology: Principles and Applications* (eds Jansonius, J. & McGregor, D. C.) 29–50 (American Association of Stratigraphic Palynologists Foundation, 1996).

70. 70.

Raine, J. I., Mildenhall, D. C. & Kennedy, E. M. *New Zealand Fossil Spores and Pollen: An Illustrated Catalogue* (GNS Science Miscellaneous Series No. 4, 4th edition, 2011);  
<http://data.gns.cri.nz/sporepollen/index.htm>

71. 71.

Prebble, J. G. Descriptions and occurrences of pollen and spores from New Zealand Cenozoic sediments. *GNS Science Internal Report* **2016**, 137 (2016).

72. 72.

Askin, R. A. in *Palaeobiology and Palaeoenvironments of Eocene Rocks, McMurdo Sound, East Antarctica Antarctic Research Series v76* (eds Stilwel, J. D. & Feldman, R. M.) 161–181 (American Geophysical Union, 2000).

73. 73.

Askin, R. A. & Raine, J. I. Oligocene and Early Miocene terrestrial palynology of the Cape Roberts Drillhole CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarct.* **7**, 493–501 (2000).

74. 74.

Truswell, E. M. Recycled Cretaceous and Tertiary pollen and spores in Antarctic marine sediments: a catalogue. *Palaeontographica Abt. B Paläophytol.* **186**, 121–174 (1983).

75. 75.

Fensome, R. A. & Williams, G. L. *The Lentin and Williams Index of Fossil Dinoflagellates* (American Association of Stratigraphic Palynologists Foundation Contribution Series 42, 2004).

76. 76.

Hannah, M. J., Wilson, G. J. & Wrenn, J. H. Oligocene and miocene marine palynomorphs from CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarct.* **7**, 503–511 (2000).

77. 77.

Hannah, M. J. The palynology of ODP site 1165, Prydz Bay, East Antarctica: a record of Miocene glacial advance and retreat. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 120–133 (2006).

78. 78.

Clowes, C. D., Hannah, M. J., Wilson, G. J. & Wrenn, J. H. Marine palynostratigraphy of the Cape Roberts Drill-holes, Victoria Land

Basin, Antarctica, with descriptions of six new species of organic-walled dinoflagellate cyst. *Mar. Micropaleontol.* **126**, 65–84 (2016).

79. 79.

Bijl, P. et al. Stratigraphic calibration of Oligocene–Miocene organic-walled dinoflagellate cysts from offshore Wilkes Land, East Antarctica, and a zonation proposal. *J. Micropalaeontol.* **37**, 105–138 (2018).

80. 80.

Benninghoff, W. S. Calculation of pollen and spores density in sediments by addition of exotic pollen in known quantities. *Pollen Spores* **6**, 332–333 (1962).

81. 81.

Harland, R. & Pudsey, C. J. Dinoflagellate cysts from sediment traps deployed in the Bellingshausen, Weddell and Scotia seas, Antarctica. *Mar. Micropaleontol.* **37**, 77–99 (1999).

82. 82.

Prebble, J. G. et al. An expanded modern dinoflagellate cyst dataset for the Southwest Pacific and Southern Hemisphere with environmental associations. *Mar. Micropaleontol.* **101**, 33–48 (2013).

83. 83.

Hartman, J. D., Bijl, P. K. & Sangiorgi, F. A review of the ecological affinities of marine organic microfossils from a Holocene record offshore of Adélie Land (East Antarctica). *J. Micropalaeontol.* **37**, 445–497 (2018).

84. 84.

Zonneveld, K. A. et al. Atlas of modern dinoflagellate cyst distribution based on 2405 data points. *Rev. Palaeobot. Palynol.* **191**, 1–197

(2013).

85. 85.

Warny, S. et al. Palynomorphs from a sediment core reveal a sudden remarkably warm Antarctica during the middle Miocene. *Geology* **37**, 955–958 (2009).

86. 86.

Sangiorgi, F. et al. Southern Ocean warming and Wilkes Land ice sheet retreat during the mid-Miocene. *Nat. Commun.* **9**, 317 (2018).

87. 87.

Niessen, F., Gebhardt, A. C., Kuhn, G., Magens, D. & Monien, D. Porosity and density of the AND-1B sediment core, McMurdo Sound region, Antarctica: Field consolidation enhanced by grounded ice. *Geosphere* **9**, 489–509 (2013).

88. 88.

Cody, R. D., Levy, R. H., Harwood, D. M. & Sadler, P. M. Thinking outside the zone: high-resolution quantitative diatom biochronology for the Antarctic Neogene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 92–121 (2008).

89. 89.

Florindo, F. et al. Paleomagnetism and biostratigraphy of sediments from Southern Ocean ODP Site 744 (southern Kerguelen Plateau): implications for early-to-middle Miocene climate in Antarctica. *Global Planet. Change* **110**, 434–454 (2013).

90. 90.

Crampton, J. S. et al. Southern Ocean phytoplankton turnover in response to stepwise Antarctic cooling over the past 15 million years. *Proc. Natl Acad. Sci. USA* **113**, 6868–6873 (2016).

91. 91.

Scherer, R., Bohaty, S. M. & Harwood, D. M. Oligocene and lower Miocene siliceous microfossil biostratigraphy of Cape Roberts Project Core CRP-2/2A, Victoria Land Basin, Antarctica. *Terra Antarct.* **7**, 417–442 (2000).

92. 92.

Taviani, M. et al. Palaeontological characterisation and analysis of the AND-2A core, ANDRILL Southern McMurdo Sound Project, Antarctica. *Terra Antarct.* **15**, 113–146 (2008).

93. 93.

Farmer, R. K. *The Application of Biostratigraphy and Paleoecology at Southern Ocean Drill Sites to Resolve Early to Middle Miocene Paleoclimatic Events* MS thesis, Univ. Nebraska-Lincoln (2011).

94. 94.

Meyers, S. R. The evaluation of eccentricity-related amplitude modulation and bundling in paleoclimate data: An inverse approach for astrochronologic testing and time scale optimization. *Paleoceanography* **30**, 1625–1640 (2015).

95. 95.

Meyers, S. R. Astrochron: An R Package for Astrochronology (2014); <http://cran.rproject.org/package=astrochron>

96. 96.

Meyers, S. R. Cyclostratigraphy and the problem of astrochronologic testing. *Earth Sci. Rev.* **190**, 190–223 (2019).

97. 97.

Laskar, J. et al. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).

98. 98.

Taner, M. T. *Attributes Revisited. Technical Report* (Rock Solid Images, 1992).

99. 99.

Billups, et al. Astronomic calibration of the late Oligocene through early Miocene geomagnetic polarity time scale. *Earth Planet. Sci. Lett.* **224**, 33–44 (2004).

100. 100.

Kochhann, K. G. et al. Eccentricity pacing of eastern equatorial Pacific carbonate dissolution cycles during the Miocene Climatic Optimum. *Paleoceanography* **31**, 1–17 (2016).

101. 101.

Suganuma, Y. et al. <sup>10</sup>Be evidence for delayed acquisition of remanent magnetization in marine sediments: Implication for a new age for the Matuyama–Brunhes boundary. *Earth Planet. Sci. Lett.* **296**, 443–450 (2010).

102. 102.

Suganuma, Y. et al. Post-depositional remanent magnetization lock-in for marine sediments deduced from <sup>10</sup>Be and paleomagnetic records through the Matuyama–Brunhes boundary. *Earth Planet. Sci. Lett.* **311**, 39–52 (2011).

103. 103.

Roberts, A. P. & Winklhofer, M. Why are geomagnetic excursions not always recorded in sediments? Constraints from post-depositional

remanent magnetization lock-in modelling. *Earth Planet. Sci. Lett.* **227**, 345–359 (2004).

104. 104.

Boger, S. D. Antarctica—before and after Gondwana. *Gondwana Res.* **19**, 335–371 (2011).

105. 105.

Siddoway, C. S. in *Antarctica: A Keystone in a Changing World* (eds Cooper, A., Raymond, C. and the 10th ISAES Editorial Team) 91–114 (The National Academic Press, USA, 2008).

106. 106.

Mukasa, S. B. & Dalziel, I. W. Marie Byrd Land, West Antarctica: Evolution of Gondwana’s Pacific margin constrained by zircon U-Pb geochronology and feldspar common-Pb isotopic compositions. *Geol. Soc. Am. Bull.* **112**, 611–627 (2000).

107. 107.

Weaver, S. D., Adams, C. J., Pankhurst, R. J. & Gibson, I. L. Granites of Edward VII Peninsula, Marie Byrd Land: anorogenic magmatism related to Antarctic-New Zealand rifting. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **83**, 281–290 (1992).

108. 108.

Korhonen, F. J., Saito, S., Brown, M., Siddoway, C. S. & Day, J. M. D. Multiple generations of granite in the Fosdick Mountains, Marie Byrd Land, West Antarctica: implications for polyphase intracrustal differentiation in a continental margin setting. *J. Petrol.* **51**, 627–670 (2010).

109. 109.

Craddock, J. et al. Precise U-Pb zircon ages and geochemistry of Jurassic granites, Ellsworth-Whitmore terrane, central Antarctica. *Geol. Soc. Am. Bull.* **129**, 118–136 (2017).

110. 110.

Pankhurst, R. J., Weaver, S. D., Bradshaw, J. D., Storey, B. C. & Ireland, T. R. Geochronology and geochemistry of pre-Jurassic superterrane in Marie Byrd Land, Antarctica. *J. Geophys. Res. Solid Earth* **103**, 2529–2547 (1998).

111. 111.

Flowerdew, M. J. et al. Combined U-Pb geochronology and Hf isotope geochemistry of detrital zircons from early Paleozoic sedimentary rocks, Ellsworth-Whitmore Mountains block, Antarctica. *Geol. Soc. Am. Bull.* **119**, 275–288 (2007).

112. 112.

Elliot, D. H. & Fanning, C. M. Detrital zircons from upper Permian and lower Triassic Victoria Group sandstones, Shackleton Glacier region, Antarctica: evidence for multiple sources along the Gondwana plate margin. *Gondwana Res.* **13**, 259–274 (2008).

113. 113.

Elliot, D. H., Fanning, C. M. & Hulett, S. R. Age provinces in the Antarctic craton: Evidence from detrital zircons in Permian strata from the Beardmore Glacier region, Antarctica. *Gondwana Res.* **28**, 152–164 (2015).

114. 114.

Goodge, J. W., Williams, I. S. & Myrow, P. Provenance of Neoproterozoic and lower Paleozoic siliciclastic rocks of the central Ross orogen, Antarctica: Detrital record of rift-, passive-, and active-margin sedimentation. *Geol. Soc. Am. Bull.* **116**, 1253–1279 (2004).

115. 115.

Paulsen, T. S. et al. Detrital mineral ages from the Ross Supergroup, Antarctica: Implications for the Queen Maud terrane and outboard sediment provenance on the Gondwana margin. *Gondwana Res.* **27**, 377–391 (2015).

116. 116.

Paulsen, T. S. et al. Correlation and Late-Stage Deformation of Liv Group Volcanics in the Ross-Delamerian Orogen, Antarctica, from New U-Pb Ages. *J. Geol.* **126**, 307–323 (2018).

117. 117.

Goodge, J. W., Fanning, C. M., Norman, M. D. & Bennett, V. C. Temporal, isotopic and spatial relations of early Paleozoic Gondwana-margin arc magmatism, central Transantarctic Mountains, Antarctica. *J. Petrol.* **53**, 2027–2065 (2012).

118. 118.

Paulsen, T. S. et al. Age and significance of ‘outboard’ high-grade metamorphics and intrusives of the Ross orogen, Antarctica. *Gondwana Res.* **24**, 349–358 (2013).

119. 119.

Rowell, A. J. et al. An active Neoproterozoic margin: evidence from the Skelton Glacier area, Transantarctic Mountains. *J. Geol. Soc. Lond.* **150**, 677–682 (1993).

120. 120.

Encarnación, J. & Grunow, A. Changing magmatic and tectonic styles along the paleo-Pacific margin of Gondwana and the onset of early Paleozoic magmatism in Antarctica. *Tectonics* **15**, 1325–1341 (1996).

121. 121.

Goodge, J. W., Hansen, V. L., Peacock, S. M., Smith, B. K. & Walker, N. W. Kinematic evolution of the Miller Range shear zone, central Transantarctic Mountains, Antarctica, and implications for Neoproterozoic to early Paleozoic tectonics of the East Antarctic margin of Gondwana. *Tectonics* **12**, 1460–1478 (1993).

122. 122.

Van Schmus, W. R., McKenna, L. W., Gonzales, D. A., Fetter, A. H. & Rowell, A. J. U-Pb geochronology of parts of the Pensacola, Thiel, and Queen Maud Mountains, Antarctica. In *The Antarctic Region: Geological Evolution and Processes* (ed. Ricci, C. A.) 187–200 (Terra Antartica Publication, 1995).

123. 123.

Stump, E. *The Ross Orogen of the Transantarctic Mountains* (Cambridge Univ. Press, 1995).

124. 124.

Martin, A. P., Price, R. C., Cooper, A. F. & McCammon, C. A. Petrogenesis of the rifted southern Victoria Land lithospheric mantle, Antarctica, inferred from petrography, geochemistry, thermobarometry and oxybarometry of peridotite and pyroxenite xenoliths from the Mount Morning eruptive centre. *J. Petrol.* **56**, 193–226 (2015).

125. 125.

Goodge, J. W., Myrow, P., Williams, I. S. & Bowring, S. A. Age and provenance of the Beardmore Group, Antarctica: constraints on Rodinia supercontinent breakup. *J. Geol.* **110**, 393–406 (2002).

126. 126.

Stump, E., Gehrels, G., Talarico, F. M. & Carosi, R. Constraints from detrital zircon geochronology on the early deformation of the Ross orogen, Transantarctic Mountains, Antarctica. In *Antarctica: A*

*Keystone in a Changing World – Online Proceedings of the 10th  
ISAES* (eds Cooper, A. K. et al.) Extended Abstract 166 (USGS Open-  
File Report 2007-1047, 2007).

127. 127.

Cooper, A. F., Maas, R., Scott, J. M. & Barber, A. J. Dating of  
volcanism and sedimentation in the Skelton Group, Transantarctic  
Mountains: implications for the Rodinia-Gondwana transition in  
southern Victoria Land, Antarctica. *Geol. Soc. Am. Bull.* **123**, 681–702  
(2011).

128. 128.

Goodge, J. W., Fanning, C. M. & Bennett, V. C. U-Pb evidence of  
~1.7 Ga crustal tectonism during the Nimrod Orogeny in the  
Transantarctic Mountains, Antarctica: implications for Proterozoic  
plate reconstructions. *Precambr. Res.* **112**, 261–288 (2001).

129. 129.

Goodge, J. W. & Fanning, C. M. Mesoarchean and Paleoproterozoic  
history of the Nimrod Complex, central Transantarctic Mountains,  
Antarctica: stratigraphic revisions and relation to the Mawson  
Continent in East Gondwana. *Precambr. Res.* **285**, 242–271 (2016).

130. 130.

Veevers, J. J. & Saeed, A. Age and composition of Antarctic bedrock  
reflected by detrital zircons, erratics, and recycled microfossils in the  
Prydz Bay–Wilkes Land–Ross Sea–Marie Byrd Land sector (70–240  
E). *Gondwana Res.* **20**, 710–738 (2011).

131. 131.

Goodge, J. W. & Fanning, C. M. 2.5 by of punctuated Earth history as  
recorded in a single rock. *Geology* **27**, 1007–1010 (1999).

132. 132.

Grindley, G. W., McGregor, V. R. & Walcott, R. I. In *Antarctic Geology: Proceedings of the First International Symposium on Antarctic Geology* (ed. Adie, R. J.) 206–219 (North Holland, 1964).

133. 133.

Laird, M. G. in *The Geology of Antarctica* (ed Tingey R. J.) 74–119 (Oxford Univ. Press, 1991).

134. 134.

Goodge, J. W. & Finn, C. A. Glimpses of East Antarctica: Aeromagnetic and satellite magnetic view from the central Transantarctic Mountains of East Antarctica. *J. Geophys. Res. Solid Earth* <https://doi.org/10.1029/2009JB006890> (2010).

135. 135.

Goodge, J. W. & Fanning, C. M. Composition and age of the East Antarctic Shield in eastern Wilkes Land determined by proxy from Oligocene-Pleistocene glaciomarine sediment and Beacon Supergroup sandstones, Antarctica. *Geol. Soc. Am. Bull.* **122**, 1135–1159 (2010).

136. 136.

Gunn, B. M. & Warren, G. Geology of Victoria Land between the Mawson and Mulock Glaciers, Antarctica. *New Zea. Geol. Bull.* **71**, 157 (1962).

137. 137.

Encarnación, J., Rowell, A. J. & Grunow, A. M. A U-Pb age for the Cambrian Taylor Formation, Antarctica: Implications for the Cambrian time scale. *J. Geol.* **107**, 497–504 (1999).

138. 138.

Wareham, C. D., Stump, E., Storey, B. C., Millar, I. L. & Riley, T. R. Petrogenesis of the Cambrian Liv Group. A bimodal volcanic rock

suite from the Ross orogen, Transantarctic Mountains. *Geol. Soc. Am. Bull.* **113**, 360–372 (2001).

139. 139.

Elliot, D. H., Larsen, D., Fanning, C. M., Fleming, T. H. & Vervoort, J. D. The Lower Jurassic Hanson Formation of the Transantarctic Mountains: implications for the Antarctic sector of the Gondwana plate margin. *Geol. Mag.* **154**, 777–803 (2016).

140. 140.

Elliot, D. H., Fanning, C. M., Isbell, J. L. & Hulett, S. R. W. The Permo-Triassic Gondwana sequence, central Transantarctic Mountains, Antarctica: Zircon geochronology, provenance, and basin evolution. *Geosphere* **13**, 155–178 (2017).

141. 141.

Elsner, M., Schöner, R., Gerdes, A. & Gaupp, R. Reconstruction of the early Mesozoic plate margin of Gondwana by U–Pb ages of detrital zircons from northern Victoria Land, Antarctica. *Geol. Soc. Lond. Spec. Publ.* **383**, 211–232 (2013).

142. 142.

Paulsen, T., Deering, C., Sliwinski, J., Bachmann, O. & Guillong, M. New detrital zircon age and trace element evidence for 1450 Ma igneous zircon sources in East Antarctica. *Precambr. Res.* **300**, 53–58 (2017).

143. 143.

Zurli, L. et al. Detrital zircons from Late Paleozoic Ice Age sequences in Victoria Land (Antarctica): New constraints on the glaciation of southern Gondwana. *Geol. Soc. Am. Bull.* (2021).

144. 144.

Welke, B. et al. Applications of detrital geochronology and thermochronology from glacial deposits to the Paleozoic and Mesozoic thermal history of the Ross Embayment, Antarctica. *Geochem. Geophys. Geosyst.* **17**, 2762–2780 (2016).

145. 145.

Vogel, M. B., Ireland, T. R. & Weaver, S. D. The multistage history of the Queen Maud Batholith, La Gorce Mountains, central Transantarctic Mountains. In *Proc. 8th Int. Symp. Antarctic Earth Sciences Wellington 1999* (eds Gamble, J. A., Skinner, D. N. B., Henrys, S. A.) 153–159 (Royal Society of New Zealand, 2002).

146. 146.

Gootee, B. & Stump, E. in *Antarctica* (eds Fütterer D. K., Damaske D., Kleinschmidt G., Miller H. & Tessensohn F.) 191–194 (Springer, 2006).

147. 147.

Barrett, P. J. in *Geology of Antarctica* (ed. Tingey, R. J.) 120–152 (Clarendon Press, 1991).

148. 148.

Ferraccioli, F., Armadillo, E., Jordan, T., Bozzo, E. & Corr, H. Aeromagnetic exploration over the East Antarctic Ice Sheet: a new view of the Wilkes Subglacial Basin. *Tectonophysics* **478**, 62–77 (2009).

149. 149.

Paxman, G. J. et al. Geology and Geomorphology of the Pensacola-Pole Basin, East Antarctica. *Geochem. Geophys. Geosyst.* **20**, 2786–2807 (2019).

150. 150.

Elliot, D. H. The Hanson Formation: a new stratigraphical unit in the Transantarctic Mountains, Antarctica. *Antarct. Sci.* **8**, 389–394 (1996).

151. 151.

Elliot, D. H. & Fleming, T. H. Occurrence and dispersal of magmas in the Jurassic Ferrar large igneous province, Antarctica. *Gondwana Res.* **7**, 223–237 (2004).

152. 152.

Burgess, S. D., Bowring, S. A., Fleming, T. H. & Elliot, D. H. High-precision geochronology links the Ferrar large igneous province with early-Jurassic ocean anoxia and biotic crisis. *Earth Planet. Sci. Lett.* **415**, 90–99 (2015).

153. 153.

Encarnación, J., Fleming, T. H., Elliot, D. H. & Eales, H. V. Synchronous emplacement of Ferrar and Karoo dolerites and the early breakup of Gondwana. *Geology* **24**, 535–538 (1996).

154. 154.

Cook, C. P. et al. Glacial erosion of East Antarctica in the Pliocene: A comparative study of multiple marine sediment provenance tracers. *Chem. Geol.* **466**, 199–218 (2017).

155. 155.

Adams, C. J. Geochronological studies of the Swanson Formation of Marie Byrd Land, West Antarctica, and correlation with northern Victoria Land, East Antarctica, and South Island, New Zealand. *N. Z. J. Geol. Geophys.* **29**, 345–358 (1986).

156. 156.

Yakymchuk, C. et al. *Anatetic Reworking and Differentiation of Continental Crust Along the Active Margin of Gondwana: A Zircon*

*Hf–O Perspective from West Antarctica* (Geological Society London, Special Publication 383, 2013); <https://doi.org/10.1144/SP383.7>

157. 157.

Yakymchuk, C. et al. Paleozoic evolution of western Marie Byrd Land, Antarctica. *Bull. Geol. Soc. Am.* **127**, 1464–1484 (2015).

158. 158.

Simões Pereira, P. et al. The geochemical and mineralogical fingerprint of West Antarctica's weak underbelly: Pine Island and Thwaites glaciers. *Chem. Geol.* **550**, 119649 (2020).

159. 159.

Adams, C. J. Geochronology of granite terranes in the Ford Ranges, Marie Byrd Land, West Antarctica. *N. Z. J. Geol. Geophys.* **30**, 51–72 (1987).

160. 160.

LeMasurier, W. E. et al. *Volcanoes of the Antarctic Plate and Southern Ocean* Vol. 48 (American Geophysical Union, 1990).

161. 161.

Licht, K. J. et al. Evidence for extending anomalous Miocene volcanism at the edge of the East Antarctic craton. *Geophys. Res. Lett.* **45**, 3009–3016 (2018).

162. 162.

Brodie, J. W. A shallow shelf around Franklin Island in the Ross Sea, Antarctica. *N. Z. J. Geol. Geophys.* **2**, 108–119 (1959).

163. 163.

Lawver, L., Lee, J., Kim, Y. & Davey, F. Flat-topped mounds in western Ross Sea: Carbonate mounds or subglacial volcanic features? *Geosphere* **8**, 645–653 (2012).

164. 164.

Di Vincenzo, G., Bracciali, L., Del Carlo, P., Panter, K. & Rocchi, S. 40Ar–39Ar dating of volcanogenic products from the AND-2A core (ANDRILL Southern McMurdo Sound Project, Antarctica): correlations with the Erebus Volcanic Province and implications for the age model of the core. *Bull. Volcanol.* **72**, 487–505 (2010).

165. 165.

Panter, K. S. et al. Melt origin across a rifted continental margin: a case for subduction-related metasomatic agents in the lithospheric source of alkaline basalt, NW Ross Sea, Antarctica. *J. Petrol.* **59**, 517–558 (2018).

166. 166.

McIntosh, W. C. 40Ar/39Ar geochronology of tephra and volcanic clasts in CRP-2A, Victoria Land Basin, Antarctica. *Terra Antarct.* **7**, 621–630 (2000).

167. 167.

LeMasurier, W. E. & Rocchi, S. Terrestrial record of post-Eocene climate history in Marie Byrd Land, West Antarctica. *Geogr. Ann., Ser. A* **87**, 51–66 (2005).

168. 168.

Rocchi, S., LeMasurier, W. E. & Di Vincenzo, G. Oligocene to Holocene erosion and glacial history in Marie Byrd Land, West Antarctica, inferred from exhumation of the Dorrel Rock intrusive complex and from volcano morphologies. *Bull. Geol. Soc. Am.* **118**, 991–1005 (2006).

169. 169.

LeMasurier, W. Shield volcanoes of Marie Byrd Land, West Antarctic rift: oceanic island similarities, continental signature, and tectonic controls. *Bull. Volcanol.* **75**, 726 (2013).

170. 170.

Behrendt, J. C. et al. Geophysical studies of the West Antarctic rift system. *Tectonics* **10**, 1257–1273 (1991).

171. 171.

McDougall, I. & Harrison, T. M. *Geochronology and Thermochronology by the 40Ar/39Ar Method* (Oxford Univ. Press, 1999).

172. 172.

Cherniak, D. J. & Watson, E. B. Pb diffusion in zircon. *Chem. Geol.* **172**, 5–24 (2001).

173. 173.

Morrison, A. D. & Reay, A. Geochemistry of Ferrar Dolerite sills and dikes at Terra Cotta Mountain, south Victoria Land, Antarctica. *Antarct. Sci.* **7**, 73–85 (1995).

174. 174.

Cox, S. C., Turnbull, I. M., Isaac, M. J., Townsend, D. B. & Smith Lyttle, B. *Geology of Southern Victoria Land, Antarctica* (Institute of Geological & Nuclear, 2012).

175. 175.

Ford, A. B. *Stratigraphy of the Layered Gabbroic Dufek Intrusion, Antarctica* (US Govt. Print. Off., 1976;  
<http://pubs.er.usgs.gov/publication/b1405D>

176. 176.

Borg, S. G., Depaolo, D. J. & Smith, B. M. Isotopic structure and tectonics of the central Transantarctic Mountains. *J. Geophys. Res. Solid Earth* **95**, 6647–6667 (1990).

177. 177.

Cox, S. C., Parkinson, D. L., Allibone, A. H. & Cooper, A. F. Isotopic character of Cambro-Ordovician plutonism, southern Victoria Land, Antarctica. *N. Z. J. Geol. Geophys.* **43**, 501–520 (2000).

178. 178.

Gunner, J. *Isotopic and Geochemical Studies of the Pre-Devonian Basement Complex, Beardmore Glacier Region, Antarctica* (Ohio State Univ. Institute of Polar Studies Report No. 41, 1976).

179. 179.

Roy, M., van de Flierdt, T., Hemming, S. R. & Goldstein, S. L.  $^{40}\text{Ar}/^{39}\text{Ar}$  ages of hornblende grains and bulk Sm/Nd isotopes of circum-Antarctic glacio-marine sediments: Implications for sediment provenance in the Southern Ocean. *Chem. Geol.* **244**, 507–519 (2007).

180. 180.

Behrendt, J. C. The aeromagnetic method as a tool to identify Cenozoic magmatism in the West Antarctic Rift System beneath the West Antarctic Ice Sheet—A review; Thiel subglacial volcano as possible source of the ash layer in the WAISCORE. *Tectonophysics* **585**, 124–136 (2013).

181. 181.

Lough, A. C. et al. Seismic detection of an active subglacial magmatic complex in Marie Byrd Land, Antarctica. *Nat. Geosci.* **6**, 1031–1035 (2013).

182. 182.

Schroeder, D. M., Blankenship, D. D., Young, D. A. & Quartini, E. Evidence for elevated and spatially variable geothermal flux beneath the West Antarctic Ice Sheet. *Proc. Natl Acad. Sci. USA* **111**, 9070–9072 (2014).

183. 183.

Ehrmann, W. U., Melles, M., Kuhn, G. & Grobe, H. Significance of clay mineral assemblages in the Antarctic Ocean. *Mar. Geol.* **107**, 249–273 (1992).

184. 184.

Fagel, N. Clay minerals, deep circulation and climate. *Proxies Late Cenozoic Paleoceanogr.* **1**, 139–184 (2007).

185. 185.

Kristoffersen, Y., Strand, K., Vorren, T., Harwood, D. & Webb, P. Pilot shallow drilling on the continental shelf, Dronning Maud Land, Antarctica. *J. Antarct. Sci.* **4**, 463–470 (2000).

186. 186.

Ehrmann, W. et al. Provenance changes between recent and glacial-time sediments in the Amundsen Sea embayment, West Antarctica: clay mineral assemblage evidence. *Antarct. Sci.* **23**, 471–486 (2011).

187. 187.

Hillenbrand, C. D., Grobe, H., Diekmann, B., Kuhn, G. & Fütterer, D. K. Distribution of clay minerals and proxies for productivity in surface sediments of the Bellingshausen and Amundsen seas (West Antarctica)—Relation to modern environmental conditions. *Mar. Geol.* **193**, 253–271 (2003).

188. 188.

Klages, J. P. et al. Temperate rainforests near the South Pole during peak Cretaceous warmth. *Nature* **580**, 81–86 (2020).

189. 189.

Zonneveld, K. A. F., Bockelmann, F. & Holzwarth, U. Selective preservation of organic-walled dinoflagellate cysts as a tool to quantify past net primary production and bottom water oxygen concentrations. *Mar. Geol.* **237**, 109–126 (2007).

190. 190.

Prebble, J. G., Hannah, M. J. & Barrett, P. J. Changing Oligocene climate recorded by palynomorphs from two glacio-eustatic sedimentary cycles, Cape Roberts Project, Victoria Land Basin, Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 58–70 (2006).

191. 191.

Kulhanek, D. K. et al. Revised chronostratigraphy of DSDP Site 270 and late Oligocene to early Miocene paleoecology of the Ross Sea sector of Antarctica. *Global Planet. Change* **178**, 46–64 (2019).

192. 192.

Feakins, S., Warny, S. & Lee, J. E. Hydrologic cycling over Antarctica during the middle Miocene warming. *Nat. Geosci.* **5**, 557–560 (2012).

193. 193.

De Santis, L., Prato, S., Brancolini, G., Lovo, M. & Torelli, L. The Eastern Ross Sea continental shelf during the Cenozoic: implications for the West Antarctic ice sheet development. *Global Planet. Change* **23**, 173–196 (1999).

194. 194.

Ford, A. B. & Barrett, P. J. Basement rocks of the south-central Ross Sea, Site 270, DSDP Leg 28. *Initial Rep. Deep Sea Drill. Proj.* **28**, 861–868 (1975).

195. 195.

Goldich, S. S., Treves, S. B., Suhr, N. H. & Stuckless, J. S. Geochemistry of the Cenozoic volcanic rocks of Ross Island and vicinity, Antarctica. *J. Geol.* **83**, 415–435 (1975).

196. 196.

Tulaczyk, S., Kamb, B., Scherer, R. P. & Engelhardt, H. F. Sedimentary processes at the base of a West Antarctic ice stream; constraints from textural and compositional properties of subglacial debris. *J. Sediment. Res.* **68**, 487–496 (1998).

197. 197.

Rosenqvist, I. T. Origin and mineralogy glacial and interglacial clays of southern Norway. *Clays Clay Miner.* **23**, 153–159 (1975).

198. 198.

Blum, J. D. & Erel, Y. Rb/ Sr isotope systematics of a granitic soil chronosequence: The importance of biotite weathering. *Geochim. Cosmochim. Acta* **61**, 3193–3204 (1997).

199. 199.

Eisenhauer, A. et al. Grain size separation and sediment mixing in Arctic Ocean sediments: evidence from the strontium isotope systematic. *Chem. Geol.* **158**, 173–188 (1999).

200. 200.

Haran, T. MODIS Mosaic of Antarctica 2008–2009 (MOA2009) Image Map, Version 1. Boulder, Colorado, USA., *NASA National*

*Snow and Ice Data Center Distributed Active Archive Center,*  
<https://doi.org/10.7265/N5KP8037> (2014).

## Acknowledgements

This research used data and samples provided by the International Ocean Discovery Program (IODP), which is sponsored by the US National Science Foundation (NSF) and participating countries under the management of Joint Oceanographic Institutions. J.W.M. was supported by a NERC DTP studentship (grant number NE/L002515/1). Neodymium and Sr isotope analysis and U–Pb dating of detrital zircons was funded through NERC UK IODP grant NE/R018219/1. Clast counts performed by L.Z., F.T. and M.P. and the participation of L.D. and F.C. was funded by the Italian National Antarctic Research Program (PNRA, Programma Nazionale Ricerche in Antartide), grant numbers PNRA18-00233, PNRA16-00016 and PNRA18-00002. R.M.M. was supported by Royal Society Te Apārangi Marsden Fund (18-VUW-089). R.M.M., J.G.P. and R.L. were supported by the New Zealand Ministry for Business Innovation and Employment grant ANTA1801. P.V. was partially funded by NERC Standard Grant NE/T001518/1. L.F.P. has been funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 792773 WAMSISE. T.E.v.P. has been funded by NERC grants NE/R018235/1 and NE/T012285/1. D.K.K. was supported by the IODP JOIDES Resolution Science Operator and National Science Foundation (grant numbers OCE-1326927 and OPP-2000995). A.E.S. and I.B. were supported by the US Science Support Program. Southern Transantarctic Mountain rock samples for Nd and Sr isotope analysis were provided by the Polar Rock Repository with support from the National Science Foundation, under Cooperative Agreement OPP-1643713. We thank B. Coles, K. Kreissig and P. Simões Pereira for technical support. We also thank the numerous scientists who collected invaluable site survey data and developed the proposals and hypotheses that ultimately led to IODP Expedition 374. Expedition 374 was conducted under Antarctic Conservation Act Permit Number: ACA 2018-027 (permit holder: Bradford Clement, JRSO, IODP, TAMU, College Station, TX 77845).

# **Author information**

## Author notes

1. Deceased: F. Talarico

## Affiliations

1. Department of Earth Science and Engineering, Imperial College London, London, UK  
J. W. Marschalek, T. van de Flierdt, M. J. Siegert & Tina van de Flierdt
2. Department of Physical, Earth and Environmental Sciences, University of Siena, Siena, Italy  
L. Zurli, F. Talarico & M. Perotti
3. Department of Earth Sciences, University College London, London, UK  
P. Vermeesch, T. E. van Peer & Tim E. van Peer
4. Department of Earth and Planetary Sciences, Birkbeck, University of London, London, UK  
A. Carter
5. Laboratoire d'Océanologie et de Géosciences, UMR 8187 CNRS/Univ Lille/ULCO, Villeneuve d'Ascq, France  
F. Beny, V. Bout-Roumazeilles & François Beny
6. Department of Earth Sciences, Marine Palynology and Paleoceanography, Utrecht University, Utrecht, the Netherlands  
F. Sangiorgi, C. Boshuis & Francesca Sangiorgi

7. Lamont-Doherty Earth Observatory of Columbia University Palisades,  
New York, NY, USA

S. R. Hemming, B. Keisling & Benjamin A. Keisling

8. British Antarctic Survey, Cambridge, UK

L. F. Pérez & C.-D. Hillenbrand

9. Geological Survey of Denmark and Greenland, Department of Marine  
Geology, Aarhus, Denmark

L. F. Pérez

10. National Institute of Oceanography and Applied Geophysics – OGS,  
Trieste, Italy

F. Colleoni, L. De Santis & Laura De Santis

11. GNS Science, Lower Hutt, New Zealand

J. G. Prebble, R. Levy & Giuseppe Cortese

12. National Oceanography Centre Southampton, University of  
Southampton Waterfront Campus, Southampton, UK

T. E. van Peer & Tim E. van Peer

13. College of Marine Science, University of South Florida, St.  
Petersburg, FL, USA

A. E. Shevenell, I. Browne, Imogen M. Browne & Amelia E.  
Shevenell

14. International Ocean Discovery Program, Texas A&M University,  
College Station, TX, USA

D. K. Kulhanek & Denise K. Kulhanek

15. Institute of Geosciences, Christian-Albrechts-University of Kiel, Kiel, Germany

D. K. Kulhanek, R. Levy & Denise K. Kulhanek

16. Antarctic Research Centre, Victoria University of Wellington, Wellington, New Zealand

R. M. McKay & Robert M. McKay

17. Department of Earth and Atmospheric Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA

D. Harwood & David M. Harwood

18. Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

N. B. Sullivan & S. R. Meyers

19. School of Earth Sciences, Ohio State University, Columbus, OH, USA

E. M. Griffith

20. Centre for Geography and Environmental Sciences, University of Exeter, Cornwall, UK

E. Gasson

21. Grantham Institute, Imperial College London, London, UK

M. J. Siegert

22. Department of Earth Sciences, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA

K. J. Licht

23. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

G. Kuhn

24. Department of Geology and Environmental Geosciences, Northern Illinois University, DeKalb, IL, USA

J. P. Dodd & Justin P. Dodd

25. Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, USA

Jeanine Ash

26. Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Bremerhaven, Germany

Oliver M. Esper

27. School of Biological & Marine Sciences, Plymouth University, Plymouth, UK

Jenny A. Gales

28. Marine Geology Research Group, Geological Survey of Japan, AIST, Tsukuba, Japan

Saki Ishino & Saiko T. Sugisaki

29. Division of Polar-Earth System Sciences, Korea Polar Research Institute, Incheon, Republic of Korea

Sookwan Kim

30. Division of Polar Paleoenvironment, Korea Polar Research Institute, Incheon, Republic of Korea

Sunghan Kim

31. Department of Geology, University of Tromsø, Tromsø, Norway

Jan Sverre Laberg

32. Department of Geosciences, University of Massachusetts, Amherst,  
Amherst, MA, USA

R. Mark Leckie

33. Marine Geology, Alfred Wegener Institute, Bremerhaven, Germany

Juliane Müller

34. Department of Geological Sciences and Environmental Studies,  
Binghamton University, State University of New York, Binghamton,  
NY, USA

Molly O. Patterson

35. Department of Geosciences, Virginia Tech, Blacksburg, VA, USA

Brian W. Romans

36. MARUM, University of Bremen, Bremen, Germany

Oscar E. Romero

37. Institute of Low Temperature Science, Hokkaido University,  
Hokkaido, Japan

Osamu Seki

38. Polar Biology Lab, National Centre for Polar and Ocean Research  
(NCPOR) Headland Sada, Vasco-da-Gama, Goa, India

Shiv M. Singh

39. Instituto de Geociencias, Universidade de Brasília, Campus  
Universitário Darcy Ribeiro, Brasília, Brazil

Isabela M. Cordeiro de Sousa

40. State Key Laboratory of Marine Geology, Tongji University, Shanghai, China

Whenshen Xiao

41. First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China

Zhifang Xiong

## **Consortia**

### **IODP Expedition 374**

- Jeanine Ash
- , François Beny
- , Imogen M. Browne
- , Giuseppe Cortese
- , Laura De Santis
- , Justin P. Dodd
- , Oliver M. Esper
- , Jenny A. Gales
- , David M. Harwood
- , Saki Ishino
- , Benjamin A. Keisling
- , Sookwan Kim
- , Sunghan Kim
- , Denise K. Kulhanek
- , Jan Sverre Laberg
- , R. Mark Leckie
- , Robert M. McKay
- , Juliane Müller
- , Molly O. Patterson
- , Brian W. Romans
- , Oscar E. Romero

- , Francesca Sangiorgi
- , Osamu Seki
- , Amelia E. Shevenell
- , Shiv M. Singh
- , Isabela M. Cordeiro de Sousa
- , Saiko T. Sugisaki
- , Tina van de Flierdt
- , Tim E. van Peer
- , Whenshen Xiao
- & Zhifang Xiong

## Contributions

J.W.M., T.v.d.F., R.M.M., L.D.S. and A.E.S. designed the research in collaboration with the entire IODP Expedition 374 science party. J.W.M. conducted the Nd and Sr isotope analyses. L.Z., F.T. and M.P. performed the clast counts. J.W.M., P.V. and A.C. produced the zircon U–Pb data. F.B. and V.B.R. collected the clay mineralogy data. F.S., J.G.P. and C.B. performed the palynological counts and interpretations. S.R.H. provided the hornblende  $^{40}\text{Ar}/^{39}\text{Ar}$  data. K.J.L. provided guidance on geochronology interpretations. L.F.P., F.C. and L.D.S. calculated the sediment volume estimate. R.L., R.M.M., T.E.v.P., D.H., D.K.K. and E.M.G. improved the shipboard age model. N.B.S. and S.R.M. conducted the astrochronological analyses. D.K.K. provided the XRF data. E.G. and B.K. helped integrate sediment provenance data with numerical modelling. I.B., G.K. and J.P.D. advised on specific technical aspects of the manuscript. J.W.M. created the figures and wrote the text with assistance from all authors and particular guidance from T.v.d.F., C.D.H., E.G. and M.J.S. All Expedition 374 scientists contributed to the collection of shipboard datasets and the interpretations of the data.

## Corresponding author

Correspondence to [J. W. Marschalek](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Patrick Blaser, Maria Fernanda Sanchez-Goni, Kenneth Miller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Age model constraints below 75 mbsf at Site U1521.](#)

From left to right are: depth (metres below sea floor), core number, core recovery (black = recovered), inclination before and after 10 and 20 mT demagnetisation (black, blue and red points, successively), and corresponding polarity interpretations (black = normal, white = reversed, grey = no interpretation). Note that the polarity interpretations have been simplified compared to those in the cruise report<sup>26</sup>, with small uncertainties related to core gaps removed. Note Site U1521 is in the Southern Hemisphere. The geomagnetic polarity timescale<sup>49</sup> is shown across the top of the plot. The orange shaded regions indicate uncertainties in our age model and the dashed line marks an alternative line of correlation for Sequence 3. The blue line indicates the age model for Sequence 2 based on our astrochronological analyses, with the light blue shading indicating the ~20 kyr uncertainty associated with the phase relationship between clast abundances and obliquity. This astrochronological anchoring agrees closely with linear interpolations between magnetostratigraphic tie points (black line).

## Extended Data Fig. 2 Selected palynological counts compared to strontium and neodymium isotope data.

Palynological data are reported as percentages (crosses) and counts/gram (circles). The blue shaded area represents Sequence 2, which is interpreted as consisting of sediments with a West Antarctic provenance. Error bars indicate a 95% confidence interval<sup>48</sup>.

## Extended Data Fig. 3 Down-core clast and clay mineral distribution.

The blue shaded area highlights Sequence 2, which is interpreted to consist of sediments with a West Antarctic provenance. **a)** Core lithology. **b)** Chronostratigraphic sequences. **c)** Clast abundance. **d)** Percentages of different clast lithologies. **e)** Ratio between dolerite and total number of clasts (red) and volcanic rocks and total number of clasts (green), with 95% confidence interval shown as pale shading<sup>48</sup>. **f)** Clay mineral abundances.

## Extended Data Fig. 4 Map of approximate $\epsilon_{\text{Nd}}$ values in rocks and offshore sediments from around the Ross Sea embayment.

Epsilon Nd values are overlain on MODIS imagery<sup>200</sup> and the BedMachine Antarctica V1 modern bed topography<sup>43,44</sup>, with the MEASUREs grounding line and ice sheet margin shown<sup>45,46</sup>. The approximate boundary between West and East Antarctic lithosphere is shown using a white dashed line<sup>47</sup>. Modern/late Holocene and terrestrial till samples are represented by circles with the same colour bar<sup>28,30,55</sup>. Although ice flow patterns have changed since their deposition, Last Glacial Maximum tills in offshore sediments are also plotted as squares to improve spatial coverage<sup>28</sup>. Individual samples and references are reported in Supplementary Table 1. The bedrock map was produced by Kriging between sample locations within a group, then masking to the outcrop area. Beacon and Ferrar Group (Fig. 1) rocks are often not differentiated in geological mapping, but are roughly equal volumetrically<sup>136</sup>, with the uppermost Beacon Supergroup formations

having a Ferrar-like isotopic signature<sup>139</sup>. We hence assume a 60% Ferrar, 40% Beacon mixture is representative.

**Extended Data Fig. 5 Kernel density estimate plots for literature measurements of rock  $\epsilon_{\text{Nd}}$  compared to measurements on fine-grained Miocene detritus from Site U1521.**

For references and a list of all the data, see Supplementary Table 1. The height of the curve indicates the density of measurements and n the total number of samples analysed. Colour scheme is identical to Fig. 1, with sediments in grey.

**Extended Data Fig. 6 Kernel density estimates for hornblende  $^{40}\text{Ar}/^{39}\text{Ar}$  ages compared to zircon U-Pb ages younger than 1500 Ma.**

The two dating methods are show in red and blue, respectively. Bold letters correspond with those in Fig. 3. The positions of major peaks and number of grains analysed are labelled in the corresponding colours. Stratigraphic position is shown in Fig. 2.

**Extended Data Fig. 7 Close up of the Site U1521 interval with a West Antarctic provenance.**

The stratigraphic log (a) is displayed alongside the percentage of reworked dinocysts (b), basalt clast fraction (c), relative abundance of smectite (d), Nd isotope data (e) and Fe/Ti ratios determined by X-ray fluorescence scanning (f).

**Extended Data Fig. 8 Correlation of Site U1521 magnetostratigraphic tie points.**

Shown are correlations between the AND-2A record<sup>11</sup>, Site U1521<sup>26</sup> and the GPTS<sup>49</sup>.

**Extended Data Table 1 Age tie points for Site U1521 below 75 mbsf**  
**Extended Data Table 2 Values used in the erosion rate calculation**

## Supplementary information

### Supplementary Information

This file contains information on the lithologies at IODP Site U1521, before summarising the rock types and tectonic history of the Ross Sea sector. We also provide a more detailed discussion of our sediment provenance datasets, plus suggested provenance interpretations for other lithological units. Additional supplementary methods are also described.

### Peer Review File

### Supplementary Table 1

Compiled Nd and Sr isotope data from literature sources are presented in this excel spreadsheet. These data were used to interpret the isotope ratios measured at Site U1521 and to create Extended Data Figures 4 and 5. References are given in a separate tab.

## Source data

### Source Data Fig. 2

### Source Data Fig. 3

## Rights and permissions

### Reprints and Permissions

## About this article

## Cite this article

Marschalek, J.W., Zurlì, L., Talarico, F. *et al.* A large West Antarctic Ice Sheet explains early Neogene sea-level amplitude. *Nature* **600**, 450–455 (2021). <https://doi.org/10.1038/s41586-021-04148-0>

- Received: 20 January 2021
- Accepted: 14 October 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04148-0>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04148-0>

- Article
- [Published: 15 December 2021](#)

# Uncovering global-scale risks from commercial chemicals in air

- [Qifan Liu](#) ORCID: [orcid.org/0000-0003-0033-5313<sup>1,2</sup>](https://orcid.org/0000-0003-0033-5313) na1,
- [Li Li](#) ORCID: [orcid.org/0000-0002-5157-7366<sup>3</sup>](https://orcid.org/0000-0002-5157-7366) na1,
- [Xianming Zhang](#) ORCID: [orcid.org/0000-0002-5301-7899<sup>1,4,5</sup>](https://orcid.org/0000-0002-5301-7899),
- [Amandeep Saini](#) ORCID: [orcid.org/0000-0003-0880-1147<sup>1</sup>](https://orcid.org/0000-0003-0880-1147),
- [Wenlong Li](#) ORCID: [orcid.org/0000-0003-3564-1708<sup>1,6</sup>](https://orcid.org/0000-0003-3564-1708),
- [Hayley Hung](#) ORCID: [orcid.org/0000-0003-0719-8948<sup>1</sup>](https://orcid.org/0000-0003-0719-8948),
- [Chunyan Hao](#) ORCID: [orcid.org/0000-0003-2885-8503<sup>4</sup>](https://orcid.org/0000-0003-2885-8503),
- [Kun Li](#) ORCID: [orcid.org/0000-0003-2970-037X<sup>1,7</sup>](https://orcid.org/0000-0003-2970-037X),
- [Patrick Lee](#) ORCID: [orcid.org/0000-0001-7438-1043<sup>1</sup>](https://orcid.org/0000-0001-7438-1043),
- [Jeremy J. B. Wentzell](#) ORCID: [orcid.org/0000-0001-9637-0142<sup>1</sup>](https://orcid.org/0000-0001-9637-0142),
- [Chunyan Huo](#) ORCID: [orcid.org/0000-0002-4226-1060<sup>1,6</sup>](https://orcid.org/0000-0002-4226-1060),
- [Shao-Meng Li](#) ORCID: [orcid.org/0000-0002-7628-6581<sup>8</sup>](https://orcid.org/0000-0002-7628-6581),
- [Tom Harner](#) ORCID: [orcid.org/0000-0001-9026-3645<sup>1</sup>](https://orcid.org/0000-0001-9026-3645) &
- [John Liggio](#) ORCID: [orcid.org/0000-0003-3683-4595<sup>1</sup>](https://orcid.org/0000-0003-3683-4595)

*Nature* volume **600**, pages 456–461 (2021)

- 1410 Accesses
- 170 Altmetric
- [Metrics details](#)

## Subjects

- [Atmospheric chemistry](#)

- [Environmental impact](#)

## Abstract

Commercial chemicals are used extensively across urban centres worldwide<sup>1</sup>, posing a potential exposure risk to 4.2 billion people<sup>2</sup>. Harmful chemicals are often assessed on the basis of their environmental persistence, accumulation in biological organisms and toxic properties, under international and national initiatives such as the Stockholm Convention<sup>3</sup>. However, existing regulatory frameworks rely largely upon knowledge of the properties of the parent chemicals, with minimal consideration given to the products of their transformation in the atmosphere. This is mainly due to a dearth of experimental data, as identifying transformation products in complex mixtures of airborne chemicals is an immense analytical challenge<sup>4</sup>. Here we develop a new framework—combining laboratory and field experiments, advanced techniques for screening suspect chemicals, and *in silico* modelling—to assess the risks of airborne chemicals, while accounting for atmospheric chemical reactions. By applying this framework to organophosphate flame retardants, as representative chemicals of emerging concern<sup>5</sup>, we find that their transformation products are globally distributed across 18 megacities, representing a previously unrecognized exposure risk for the world's urban populations. More importantly, individual transformation products can be more toxic and up to an order-of-magnitude more persistent than the parent chemicals, such that the overall risks associated with the mixture of transformation products are also higher than those of the parent flame retardants. Together our results highlight the need to consider atmospheric transformations when assessing the risks of commercial chemicals.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

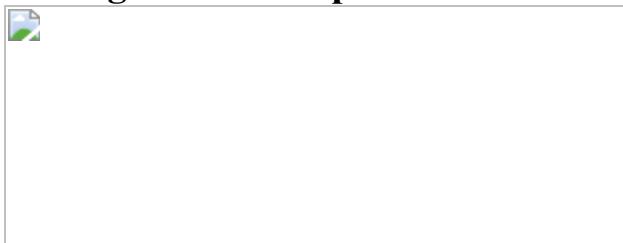
[Rent or Buy](#)

All prices are NET prices.

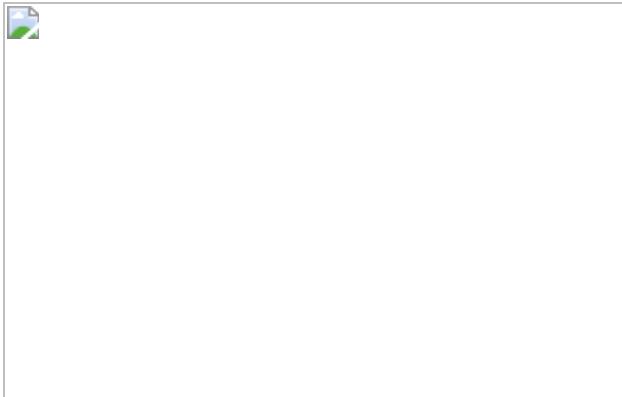
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

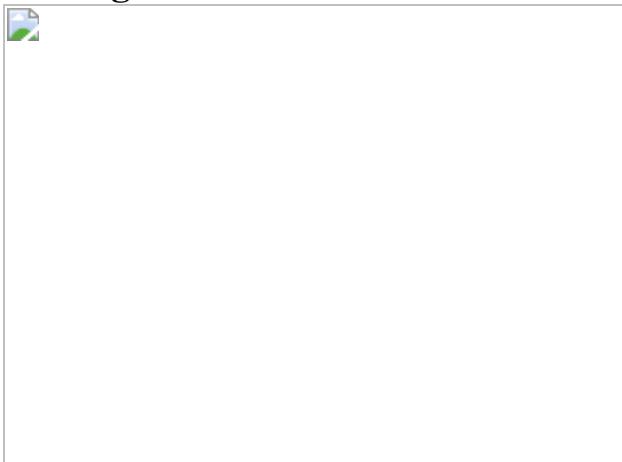
**Fig. 1: Proposed framework for the improved risk assessment of CECs, including their atmospheric transformations.**



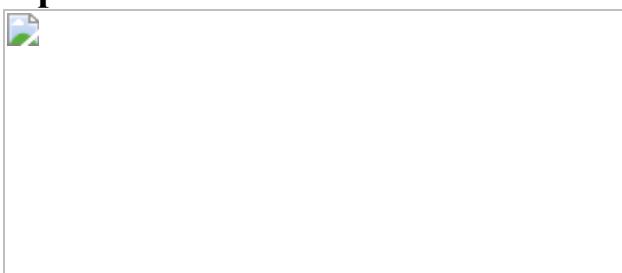
**Fig. 2: Global distribution of atmospheric OPFR transformation products.**



**Fig. 3: Sum of concentrations of OPFR transformation products found in 18 megacities.**



**Fig. 4: Persistence, bioaccumulation, toxicity and risks of individual OPFR transformation products and their mixture relative to the parent compounds.**



## Data availability

The full data set shown in the figures and tables is publicly available at <https://doi.org/10.5281/zenodo.5251825>. [Source data](#) are provided with this paper.

# Code availability

The RAIDAR model is implemented in a user-friendly online platform named the Exposure and Safety Estimation (EAS-E) Suite, which is free for online use at [www.eas-e-suite.com](http://www.eas-e-suite.com).

## References

1. 1.

Johnson, A. C., Jin, X., Nakada, N. & Sumpter, J. P. Learning from the past and considering the future of chemicals in the environment. *Science* **367**, 384–387 (2020).

2. 2.

United Nations. *World Urbanization Prospects*.  
<https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf> (UN, 2018).

3. 3.

United Nations. *Stockholm Convention*. <http://www.pops.int/> (UN, 2004).

4. 4.

Escher, B. I., Stapleton, H. M. & Schymanski, E. L. Tracking complex mixtures of chemicals in our changing environment. *Science* **367**, 388–392 (2020).

5. 5.

De Boer, J. & Stapleton, H. M. Toward fire safety without chemical risk. *Science* **364**, 231–232 (2019).

6. 6.

Wang, Z. et al. We need a global science-policy body on chemicals and waste. *Science* **371**, 774–776 (2021).

7. 7.

United States. *Toxic Substances Control Act*. <https://www.epa.gov/tsca-inventory> (US, 1976).

8. 8.

European Union. *Registration Evaluation, Authorization and Restriction of Chemicals*.

<https://echa.europa.eu/regulations/reach/understanding-reach> (EU, 2007).

9. 9.

Environment and Climate Change Canada. *Science Approach Document: Ecological Risk Classification of Organic Substances*. [https://www.ec.gc.ca/ese-ees/A96E2E98-2A04-40C8-9EDC-08A6DFF235F7/CMP3%20ERC\\_EN.pdf](https://www.ec.gc.ca/ese-ees/A96E2E98-2A04-40C8-9EDC-08A6DFF235F7/CMP3%20ERC_EN.pdf) (ECCC, 2016).

10. 10.

Jones, K. C. Persistent organic pollutants (POPs) and related chemicals in the global environment: some personal reflections. *Environ. Sci. Technol.* **55**, 9400–9412 (2021).

11. 11.

Saini, A. et al. GAPS-megacities: a new global platform for investigating persistent organic pollutants and chemicals of emerging concern in urban air. *Environ. Pollut.* **267**, 115416 (2020).

12. 12.

Salamova, A., Ma, Y., Venier, M. & Hites, R. A. High levels of organophosphate flame retardants in the Great Lakes atmosphere. *Environ. Sci. Technol. Lett.* **1**, 8–14 (2013).

13. 13.

Molina, M. J. & Rowland, F. S. Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. *Nature* **249**, 810–812 (1974).

14. 14.

Washington, J. W. et al. Nontargeted mass-spectral detection of chloroperfluoropolyether carboxylates in New Jersey soils. *Science* **368**, 1103–1107 (2020).

15. 15.

Hollender, J., Schymanski, E. L., Singer, H. P. & Ferguson, P. L. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ. Sci. Technol.* **51**, 11505–11512 (2017).

16. 16.

Hites, R. A. & Jobst, K. J. Is nontargeted screening reproducible? *Environ. Sci. Technol.* **52**, 11975–11976 (2018).

17. 17.

Blum, A. et al. Organophosphate ester flame retardants: are they a regrettable substitution for polybrominated diphenyl ethers? *Environ. Sci. Technol. Lett.* **6**, 638–649 (2019).

18. 18.

Fenner, K., Kooijman, C., Scheringer, M. & Hungerbühler, K. Including transformation products into the risk assessment for chemicals: the case of nonylphenol ethoxylate usage in Switzerland. *Environ. Sci. Technol.* **36**, 1147–1154 (2002).

19. 19.

Lienert, J., Güdel, K. & Escher, B. I. Screening method for ecotoxicological hazard assessment of 42 pharmaceuticals considering human metabolism and excretory routes. *Environ. Sci. Technol.* **41**, 4471–4478 (2007).

20. 20.

Peng, Z. & Jimenez, J. L. Radical chemistry in oxidation flow reactors for atmospheric chemistry research. *Chem. Soc. Rev.* **49**, 2570–2616 (2020).

21. 21.

Lopez-Hilfiker, F. D. et al. An extractive electrospray ionization time-of-flight mass spectrometer (EESI-TOF) for online measurement of atmospheric aerosol particles. *Atmos. Meas. Tech.* **12**, 4867–4886 (2019).

22. 22.

Zheng, Z., Peters, G. M., Arp, H. P. H. & Andersson, P. L. Combining in silico tools with multicriteria analysis for alternatives assessment of hazardous chemicals: a case study of decabromodiphenyl ether alternatives. *Environ. Sci. Technol.* **53**, 6341–6351 (2019).

23. 23.

Harrison, R. M. Urban atmospheric chemistry: a very special case for study. *NPJ Clim. Atmos. Sci.* **1**, 1–5 (2018).

24. 24.

Jimenez, J. L. et al. Evolution of organic aerosols in the atmosphere. *Science* **326**, 1525–1529 (2009).

25. 25.

Saini, A. et al. Flame retardants in urban air: a case study in Toronto targeting distinct source sectors. *Environ. Pollut.* **247**, 89–97 (2019).

26. 26.

Kelly, B. C., Ikonomou, M. G., Blair, J. D., Morin, A. E. & Gobas, F. A. Food web-specific biomagnification of persistent organic pollutants. *Science* **317**, 236–239 (2007).

27. 27.

Neumann, M. & Schliebner, I. *Protecting the Sources of our Drinking Water: the Criteria for Identifying Persistent, Mobile and Toxic (PMT) Substances and Very Persistent and Very Mobile (vPvM) Substances under EU Regulation REACH (EC) No 1907/2006* (German Environment Agency, 2019).

28. 28.

Kalgutkar, A. S. et al. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.* **6**, 161–225 (2005).

29. 29.

Zhang, Q., Yu, C., Fu, L., Gu, S. & Wang, C. New insights in the endocrine disrupting effects of three primary metabolites of organophosphate flame retardants. *Environ. Sci. Technol.* **54**, 4465–4474 (2020).

30. 30.

Tian, Z. et al. A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon. *Science* **371**, 185–189 (2021).

31. 31.

Vasiljevic, T. & Harner, T. Bisphenol A and its analogues in outdoor and indoor air: properties, sources and global levels. *Sci. Total Environ.* **789**, 148013 (2021).

32. 32.

Vethaak, A. D. & Legler, J. Microplastics and human health. *Science* **371**, 672–674 (2021).

33. 33.

Su, H. et al. Persistent, bioaccumulative, and toxic properties of liquid crystal monomers and their detection in indoor residential dust. *Proc. Natl Acad. Sci. USA* **116**, 26450–26458 (2019).

34. 34.

Liu, Q. et al. Experimental study of OH-initiated heterogeneous oxidation of organophosphate flame retardants: kinetics, mechanism, and toxicity. *Environ. Sci. Technol.* **53**, 14398–14408 (2019).

35. 35.

Liu, Q., Liggio, J., Li, K., Lee, P. & Li, S. M. Understanding the impact of relative humidity and coexisting soluble iron on the OH-initiated heterogeneous oxidation of organophosphate flame retardants. *Environ. Sci. Technol.* **53**, 6794–6803 (2019).

36. 36.

Liu, Q. et al. Atmospheric OH oxidation chemistry of particulate liquid crystal monomers: an emerging persistent organic pollutant in air. *Environ. Sci. Technol. Lett.* **7**, 646–652 (2020).

37. 37.

Liu, Y. & Sander, S. P. Rate constant for the OH + CO reaction at low temperatures. *J. Phys. Chem. A* **119**, 10060–10066 (2015).

38. 38.

Qi, L. et al. Organic aerosol source apportionment in Zurich using an extractive electrospray ionization time-of-flight mass spectrometer (EESI-TOF-MS)—part 2: biomass burning influences in winter. *Atmos. Chem. Phys.* **19**, 8037–8062 (2019).

39. 39.

Zhang, X., Saini, A., Hao, C. & Harner, T. Passive air sampling and nontargeted analysis for screening POP-like chemicals in the atmosphere: opportunities and challenges. *TrAC. Trends Analyt. Chem.* **132**, 116052 (2020).

40. 40.

Krapf, M. et al. Labile peroxides in secondary organic aerosol. *Chem* **1**, 603–616 (2016).

41. 41.

Buchholz, A. et al. Deconvolution of FIGAERO–CIMS thermal desorption profiles using positive matrix factorisation to identify chemical and physical processes during particle evaporation. *Atmos. Chem. Phys.* **20**, 7693–7716 (2020).

42. 42.

Arnot, J. A., Mackay, D., Webster, E. & Southwood, J. M. Screening level risk assessment model for chemical fate and effects in the environment. *Environ. Sci. Technol.* **40**, 2316–2323 (2006).

43. 43.

Arnot, J. A. & Mackay, D. Policies for chemical hazard and risk priority setting: can persistence, bioaccumulation, toxicity, and quantity information be combined? *Environ. Sci. Technol.* **42**, 4648–4654 (2008).

44. 44.

Arnot, J. A., Brown, T. N., Wania, F., Breivik, K. & McLachlan, M. S. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. *Environ. Health Perspect.* **120**, 1565–1570 (2012).

45. 45.

Ring, C. L. et al. Consensus modeling of median chemical intake for the US population based on predictions of exposure pathways. *Environ. Sci. Technol.* **53**, 719–732 (2018).

46. 46.

Arnot, J. A., Toose, L. & Armitage, J. *Generation of Physical-Chemical Property Data and the Application of Models for Estimating Fate and Transport and Exposure and Risk Potential for Organic Substances on the Canadian DSL* (Technical Report for Environment and Climate Change Canada, 2018).

47. 47.

Arnot, J. A. & Armitage, J. M. *Parameterization and Application of the RAIDAR Model to Aid in the Prioritization and Assessment of Chemical Substances* (Technical Report for Health Canada, 2013).

48. 48.

Arnot, J. A. & Mackay, D. *Risk Prioritization for a Subset of Domestic Substances List Chemicals using the RAIDAR model (CEMC Report No. 200703)* (Technical Report for Environment and Climate Change Canada, 2007).

49. 49.

Klasmeier, et al. Application of multimedia models for screening assessment of long-range transport potential and overall persistence. *Environ. Sci. Technol.* **40**, 53–60 (2006).

50. 50.

Webster, E., Mackay, D. & Wania, F. Evaluating environmental persistence. *Environ. Toxicol. Chem.* **17**, 2148–2158 (1998).

51. 51.

Tang, B. et al. Bioconcentration and biotransformation of organophosphorus flame retardants (PFRs) in common carp (*Cyprinus carpio*). *Environ. Int.* **126**, 512–522 (2019).

52. 52.

Bekele, T. G., Zhao, H., Wang, Y., Jiang, J. & Tan, F. Measurement and prediction of bioconcentration factors of organophosphate flame retardants in common carp (*Cyprinus carpio*). *Ecotoxicol. Environ. Saf.* **166**, 270–276 (2018).

53. 53.

Wang, G. et al. Bioaccumulation mechanism of organophosphate esters in adult zebrafish (*Danio rerio*). *Environ. Pollut.* **229**, 177–187 (2017).

54. 54.

Muir, D., Yarechewski, A. & Grift, N. Environmental dynamics of phosphate esters. III. comparison of the bioconcentration of four triaryl phosphates by fish. *Chemosphere* **12**, 155–166 (1983).

55. 55.

Mansouri, K., Grulke, C. M., Judson, R. S. & Williams, A. J. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J. Cheminform.* **10**, 1–19 (2018).

56. 56.

Trapp, S. & Horobin, R. W. A predictive model for the selective accumulation of chemicals in tumor cells. *Eur. Biophys. J.* **34**, 959–966 (2005).

57. 57.

Mansouri, K. et al. Open-source QSAR models for pKa prediction using multiple machine learning approaches. *J. Cheminform.* **11**, 1–20 (2019).

58. 58.

Brown, T. N., Arnot, J. A. & Wania, F. Iterative fragment selection: a group contribution approach to predicting fish biotransformation half-lives. *Environ. Sci. Technol.* **46**, 8253–8260 (2012).

59. 59.

Arnot JA, Gouin T, Mackay D. *Practical Methods for Estimating Environmental Biodegradation Rates* (Report for Environment Canada, Canadian Environmental Modelling Network, 2005).

60. 60.

Martin, T. *User's Guide for T.E.S.T. (version 5.1) Toxicity Estimation Software Tool: A Program to Estimate Toxicity from Molecular Structure* (Sustainable Technology Division, National Risk Management Research Laboratory, US Environmental Protection Agency, Cincinnati, OH, 2020).

61. 61.

van Leeuwen, C. J. & Vermeire, T. G. *Risk Assessment of Chemicals: An Introduction* (Springer, 2007).

62. 62.

Escher, B. I. et al. Environmental toxicology and risk assessment of pharmaceuticals from hospital wastewater. *Water Res.* **45**, 75–92 (2011).

63. 63.

Escher, B. I. & Fenner, K. Recent advances in environmental risk assessment of transformation products. *Environ. Sci. Technol.* **45**, 3835–3847 (2011).

## Acknowledgements

We thank J. A. Arnot for providing the codes for the latest version of the RAIDAR model. We acknowledge funding support from the Air Pollution programme of Environment and Climate Change Canada (ECCC). This work was also partially funded by the Chemicals Management Plan (CMP). It does not reflect any regulatory conclusions for any substances mentioned.

## Author information

### Author notes

1. These authors contributed equally: Qifan Liu, Li Li

### Affiliations

1. Air Quality Processes Research Section, Environment and Climate Change Canada, Toronto, Ontario, Canada

Qifan Liu, Xianming Zhang, Amandeep Saini, Wenlong Li, Hayley Hung, Kun Li, Patrick Lee, Jeremy J. B. Wentzell, Chunyan Huo, Tom Harner & John Liggio

2. Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei, China

Qifan Liu

3. School of Public Health, University of Nevada Reno, Reno, NV, USA

Li Li

4. Laboratory Services Branch, Ontario Ministry of the Environment, Conservation and Parks, Toronto, Ontario, Canada

Xianming Zhang & Chunyan Hao

5. Department of Chemistry and Biochemistry, Concordia University, Montreal, Quebec, Canada

Xianming Zhang

6. International Joint Research Center for Persistent Toxic Substances (IJRC-PTS), State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin, China

Wenlong Li & Chunyan Huo

7. Laboratory of Atmospheric Chemistry, Paul Scherrer Institute (PSI), Villigen, Switzerland

Kun Li

8. State Key Joint Laboratory of Environmental Simulation and Pollution Control, Beijing Innovation Center for Engineering Science and Advanced Technology, College of Environmental Sciences and Engineering, Peking University, Beijing, China

Shao-Meng Li

## Contributions

Q.L., J.L. and T.H. designed the research. A.S. and T.H. led the GAPS–MC programme. A.S. carried out the target analysis experiments for parent OPFRs. X.Z. and Chunyan Hao carried out the NTA experiments for GAPS–MC samples. Q.L., X.Z. and W.L. analysed the non-target mass spectra data. Q.L. carried out the OFR experiments. L.L. performed the *in silico* modelling. Q.L., L.L., J.L., X.Z., A.S., T.H., H.H., Chunyan Hao, W.L., K.L., P.L., J.J.B.W., Chunyan Huo and S.-M.L. contributed to the scientific discussions. Q.L., L.L., J.L., A.S. and X.Z. wrote the manuscript.

## Corresponding authors

Correspondence to [Tom Harner](#) or [John Liggio](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Kevin Jones and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Chemical structures of OPFR parent compounds and transformation products studied here.](#)

**a**, Nine OPFR parent compounds. **b**, Ten OPFR transformation products identified in GAPS–MC field samples. TCEP-1 and TCPP-25 (see Supplementary Table 1) are difficult to distinguish from each other as they have identical chemical formulae. However, given that TCEP-1 and TCPP-25 are formed through one-step and two-step photooxidation reactions, respectively, and given that the oxidation timescales associated with urban regions are short<sup>23</sup>, TCEP-1 is likely to be the dominant product.

### [Extended Data Fig. 2 Generalized mechanism for the photooxidation of atmospheric OPFRs.](#)

Photooxidation reactions proceed predominantly through three main channels. **a**, OH addition to the substituents attached to the phosphate centre (channel 1). **b**, OH addition to the phosphate centre (channel 2). **c**, Photodecomposition of early-generation products (channel 3). Generalized examples for chlorinated and non-chlorinated OPFRs are shown here.

Detailed mechanisms for TCPP (a chlorinated OPFR) and EHDP (a non-chlorinated OPFR) are shown in Supplementary Figs. [1](#), [2](#).

**Extended Data Fig. 3 Mass chromatograms and spectra for six OPFR transformation products in OFR laboratory samples and GAPS–MC field samples.**

This information is used in the identification of: **a**, TCEP-1; **b**, TCEP-10; **c**, TCEP-21; **d**, TCPP-9; **e**, TCPP-21; and **f**, TCPP-38. Detailed product information—including chemical formulae, retention times, measured *m/z* values, isotopic ratios, detection frequencies and concentrations—is summarized in Supplementary Tables [2](#), [4](#).

[Source data](#)

**Extended Data Fig. 4 Mass chromatograms and spectra of four OPFR transformation products in OFR laboratory samples and GAPS–MC field samples.**

This information is used in the identification of: **a**, TDCPP-14; **b**, TBEP-36; **c**, TPhP-6; and **d**, TPhP-8. Detailed product information—including chemical formulae, retention times, measured *m/z* values, isotopic ratios, detection frequencies and concentrations—is summarized in Supplementary Tables [2](#), [4](#).

[Source data](#)

**Extended Data Fig. 5 Seasonal variation in  $R_{\text{signal}}$  in Toronto, Canada, as a measure of photochemical production of OPFR products.**

$R_{\text{signal}}$  is the ratio of the signal intensity of an OPFR product to the signal intensity of the corresponding parent OPFR. A higher  $R_{\text{signal}}$  is indicative of increased photooxidation in the summer months. Samples were collected

during summer (August to September 2016) and winter (December 2016 to January 2017).

[Source data](#)

**Extended Data Fig. 6 Persistence results for OPFRs and their transformation products.**

**a**, Modelled overall persistence for three chlorinated OPFRs (TCEP, TCPP and TDCCP) and their products. **b**, Modelled overall persistence for six non-chlorinated OPFRs (TBEP, TPhP, EHDP, TCP, TEHP and DPhP) and their products. **c**, Relative persistence for three chlorinated OPFRs and their transformation products (parent compound = 1). **d**, Relative persistence for six non-chlorinated OPFRs and their transformation products. Dark red, light red, dark blue and light blue represent chlorinated parent OPFRs, chlorinated OPFR products, non-chlorinated parent OPFRs, and non-chlorinated OPFR products, respectively. Note that a higher overall persistence indicates a higher persistence in the multimedia environment. Compounds marked with asterisks are those identified in megacity field samples.

[Source data](#)

**Extended Data Fig. 7 Octanol-air ( $\log K_{OA}$ ) and octanol-water ( $\log K_{OW}$ ) partition coefficients for OPFRs and their transformation products at a pH of 7.**

Dark red, light red, dark blue and light blue represent chlorinated parent OPFRs, chlorinated OPFR products, non-chlorinated parent OPFRs and non-chlorinated OPFR products, respectively.

[Source data](#)

**Extended Data Fig. 8 Bioaccumulation of OPFRs and their transformation products in aquatic organisms.**

**a**, Modelled bioconcentration factor (BCF) for three chlorinated OPFRs (TCEP, TCPP and TDCPP) and their products. **b**, Modelled bioconcentration factor for six non-chlorinated OPFRs (TBEP, TPhP, EHDP, TCP, TEHP and DPhP) and their products. **c**, Relative bioaccumulation for three chlorinated OPFRs and their transformation products (parent compound = 1). **d**, Relative bioaccumulation for six non-chlorinated OPFRs and their transformation products. Dark red, light red, dark blue and light blue represent chlorinated parent OPFRs, chlorinated OPFR products, non-chlorinated parent OPFRs and non-chlorinated OPFR products, respectively. Note that a higher bioconcentration factor indicates a higher potential for bioaccumulation in aquatic organisms. Compounds marked with asterisks are those identified in megacity field samples.

[Source data](#)

**Extended Data Fig. 9 Toxicity of OPFRs and their transformation products towards aquatic organisms.**

**a**, Modelled fathead minnow LC<sub>50</sub> values for three chlorinated OPFRs (TCEP, TCPP and TDCPP) and their products. **b**, Modelled fathead minnow LC<sub>50</sub> values for six non-chlorinated OPFRs (TBEP, TPhP, EHDP, TCP, TEHP and DPhP) and their products. **c**, Relative toxicity for three chlorinated OPFRs and their transformation products (parent compound = 1). **d**, Relative toxicity for six non-chlorinated OPFRs and their transformation products. Dark red, light red, dark blue and light blue represent chlorinated parent OPFRs, chlorinated OPFR products, non-chlorinated parent OPFRs and non-chlorinated OPFR products, respectively. Note that a lower fathead minnow LC<sub>50</sub> indicates a higher toxicity. Compounds marked with asterisks are those identified in megacity field samples.

[Source data](#)

**Extended Data Table 1 Passive air-sampling details for the 18 megacities**

# **Supplementary information**

## **Supplementary Information**

This file contains Supplementary Methods and Supplementary Discussion

## **Peer Review File**

## **Source data**

### **Source Data Fig. 2**

### **Source Data Fig. 3**

### **Source Data Fig. 4**

### **Source Data Extended Data Fig. 3**

### **Source Data Extended Data Fig. 4**

### **Source Data Extended Data Fig. 5**

### **Source Data Extended Data Fig. 6**

### **Source Data Extended Data Fig. 7**

### **Source Data Extended Data Fig. 8**

### **Source Data Extended Data Fig. 9**

## **Rights and permissions**

### **Reprints and Permissions**

# About this article

## Cite this article

Liu, Q., Li, L., Zhang, X. *et al.* Uncovering global-scale risks from commercial chemicals in air. *Nature* **600**, 456–461 (2021).  
<https://doi.org/10.1038/s41586-021-04134-6>

- Received: 04 February 2021
- Accepted: 13 October 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04134-6>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04134-6>

- Article
- [Published: 15 December 2021](#)

# Deep-mantle krypton reveals Earth's early accretion of carbonaceous matter

- [Sandrine Péron](#) ORCID: [orcid.org/0000-0003-2198-7935](https://orcid.org/0000-0003-2198-7935)<sup>1</sup>,
- [Sujoy Mukhopadhyay](#)<sup>1</sup>,
- [Mark D. Kurz](#)<sup>2</sup> &
- [David W. Graham](#) ORCID: [orcid.org/0000-0002-7411-1905](https://orcid.org/0000-0002-7411-1905)<sup>3</sup>

*Nature* volume 600, pages 462–467 (2021)

- 584 Accesses
- 98 Altmetric
- [Metrics details](#)

## Subjects

- [Geochemistry](#)
- [Inner planets](#)
- [Meteoritics](#)

## Abstract

Establishing when, and from where, carbon, nitrogen and water were delivered to Earth is a fundamental objective in understanding the origin of habitable planets such as Earth. Yet, volatile delivery to Earth remains

controversial<sup>1,2,3,4,5</sup>. Krypton isotopes provide insights on volatile delivery owing to their substantial isotopic variations among sources<sup>6,7,8,9,10</sup>, although pervasive atmospheric contamination has hampered analytical efforts. Here we present the full suite of krypton isotopes from the deep mantle of the Galápagos and Iceland plumes, which have the most primitive helium, neon and tungsten isotopic compositions<sup>11,12,13,14,15,16</sup>. Except for  $^{86}\text{Kr}$ , the krypton isotopic compositions are similar to a mixture of chondritic and atmospheric krypton. These results suggest early accretion of carbonaceous material by proto-Earth and rule out any combination of hydrodynamic loss with outgassing of the deep or shallow mantle to explain atmospheric noble gases. Unexpectedly, the deep-mantle sources have a deficit in the neutron-rich  $^{86}\text{Kr}$  relative to the average composition of carbonaceous meteorites, which suggests a nucleosynthetic anomaly. Although the relative depletion of neutron-rich isotopes on Earth compared with carbonaceous meteorites has been documented for a range of refractory elements<sup>1,17,18</sup>, our observations suggest such a depletion for a volatile element. This finding indicates that accretion of volatile and refractory elements occurred simultaneously, with krypton recording concomitant accretion of non-solar volatiles from more than one type of material, possibly including outer Solar System planetesimals.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

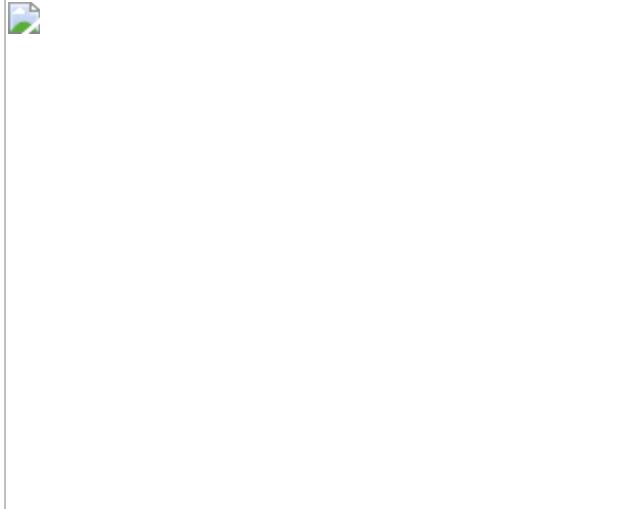
[Rent or Buy](#)

All prices are NET prices.

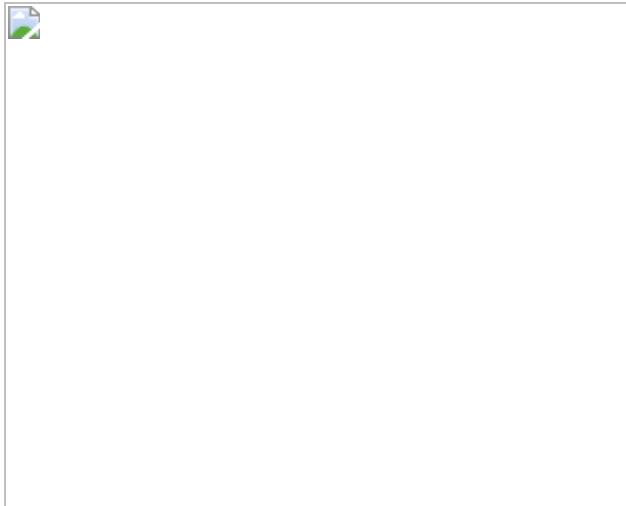
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

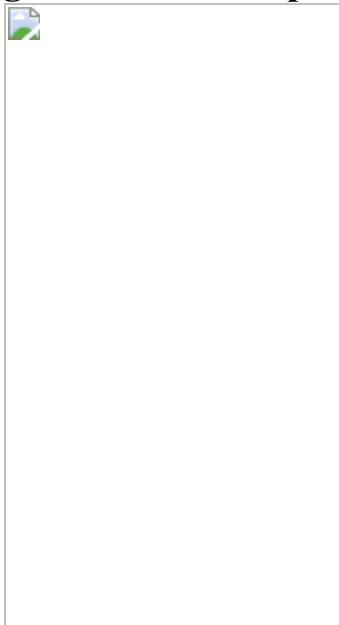
**Fig. 1: Krypton isotopic patterns of Galápagos and Iceland samples.**



**Fig. 2: Krypton isotopic compositions of Galápagos and Iceland samples.**



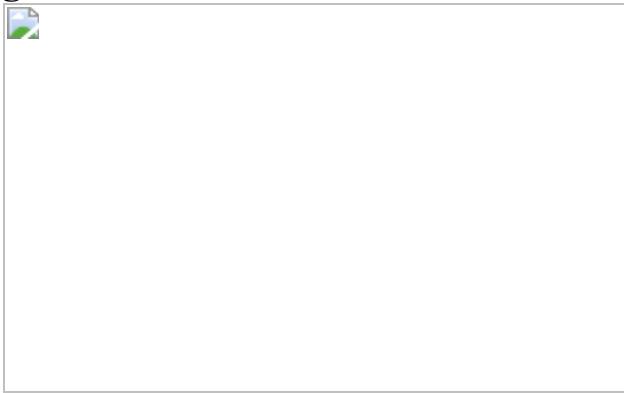
**Fig. 3: Xenon isotope composition of Galápagos and Iceland samples.**



**Fig. 4: Mixing proportions in the plume sources.**



**Fig. 5: Estimated  $^{86}\text{Kr}/^{84}\text{Kr}$  ratio based on different mixing scenarios.**



## Data availability

The geochemical data that support the findings of this study are archived on EarthChem at <https://ecl.earthchem.org/view.php?id=2065>. Source data are provided with this paper.

## References

1. 1.

Fischer-Gödde, M. et al. Ruthenium isotope vestige of Earth's pre-late-veneer mantle preserved in Archaean rocks. *Nature* **579**, 240–244

(2020).

2. 2.

Albarède, F. Volatile accretion history of the terrestrial planets and dynamic implications. *Nature* **461**, 1227–1233 (2009).

3. 3.

Mukhopadhyay, S. & Parai, R. Noble gases: a record of Earth's evolution and mantle dynamics. *Annu. Rev. Earth Planet. Sci.* **47**, 389–419 (2019).

4. 4.

Moreira, M. Noble gas constraints on the origin and evolution of Earth's volatiles. *Geochem. Perspect.* **2**, 229–403 (2013).

5. 5.

Clay, P. L. et al. Halogens in chondritic meteorites and terrestrial accretion. *Nature* **551**, 614–618 (2017).

6. 6.

Ozima, M. & Zashu, S. Solar-type Ne in Zaire cubic diamonds. *Geochim. Cosmochim. Acta* **52**, 19–25 (1988).

7. 7.

Ozima, M. & Zashu, S. Noble gas state of the ancient mantle as deduced from noble gases in coated diamonds. *Earth Planet. Sci. Lett.* **105**, 13–27 (1991).

8. 8.

Péron, S. & Moreira, M. Onset of volatile recycling into the mantle determined by xenon anomalies. *Geochem. Perspect. Lett.* **9**, 21–25 (2018).

9. 9.

Holland, G., Cassidy, M. & Ballentine, C. J. Meteorite Kr in Earth's mantle suggests a late accretionary source for the atmosphere. *Science* **326**, 1522–1525 (2009).

10. 10.

Broadley, M. W. et al. Identification of chondritic krypton and xenon in Yellowstone gases and the timing of terrestrial volatile accretion. *Proc. Natl Acad. Sci. USA* **117**, 13997–14004 (2020).

11. 11.

Péron, S. et al. Neon isotopic composition of the mantle constrained by single vesicle analyses. *Earth Planet. Sci. Lett.* **449**, 145–154 (2016).

12. 12.

Mundl-Petermeier, A. et al. Anomalous  $^{182}\text{W}$  in high  $^3\text{He}/^4\text{He}$  ocean island basalts: fingerprints of Earth's core? *Geochim. Cosmochim. Acta* **271**, 194–211 (2020).

13. 13.

Kurz, M. D., Curtice, J., Fornari, D., Geist, D. & Moreira, M. Primitive neon from the center of the Galapagos hotspot. *Earth Planet. Sci. Lett.* **286**, 23–34 (2009).

14. 14.

Mukhopadhyay, S. Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* **486**, 101–104 (2012).

15. 15.

Trieloff, M., Kunz, J., Clague, D. A., Harrison, D. & Allègre, C. J. The nature of pristine noble gases in mantle plumes. *Science* **288**, 1036–1038 (2000).

16. 16.

Harrison, D., Burnard, P. & Turner, G. Noble gas behaviour and composition in the mantle: constraints from the Iceland Plume. *Earth Planet. Sci. Lett.* **171**, 199–207 (1999).

17. 17.

Render, J., Fischer-Gödde, M., Burkhardt, C. & Kleine, T. The cosmic molybdenum–neodymium isotope correlation and the building material of the Earth. *Geochem. Perspect. Lett.* **3**, 170–178 (2017).

18. 18.

Akram, W., Schönbächler, M., Bisterzo, S. & Gallino, R. Zirconium isotope evidence for the heterogeneous distribution of s-process materials in the Solar System. *Geochim. Cosmochim. Acta* **165**, 484–500 (2015).

19. 19.

Hirschmann, M. M. & Dasgupta, R. The H/C ratios of Earth’s near-surface and deep reservoirs, and consequences for deep Earth volatile cycles. *Chem. Geol.* **262**, 4–16 (2009).

20. 20.

Marty, B. The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).

21. 21.

Marty, B. et al. Xenon isotopes in 67P/Churyumov–Gerasimenko show that comets contributed to Earth’s atmosphere. *Science* **356**, 1069–1072 (2017).

22. 22.

Harper, C. L. & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273**, 1814–1818 (1996).

23. 23.

O'Brien, D. P., Walsh, K. J., Morbidelli, A., Raymond, S. N. & Mandell, A. M. Water delivery and giant impacts in the 'grand tack' scenario. *Icarus* **239**, 74–84 (2014).

24. 24.

Meshik, A., Hohenberg, C., Pravdivtseva, O. & Burnett, D. Heavy noble gases in solar wind delivered by Genesis mission. *Geochim. Cosmochim. Acta* **127**, 326–347 (2014).

25. 25.

Busemann, H., Baur, H. & Wieler, R. Primordial noble gases in 'Phase Q' in carbonaceous and ordinary chondrites studied by closed-system stepped etching. *Meteorit. Planet. Sci.* **35**, 949–973 (2000).

26. 26.

Williams, C. D. & Mukhopadhyay, S. Capture of nebular gases during Earth's accretion is preserved in deep-mantle neon. *Nature* **565**, 78–81 (2019).

27. 27.

Yokochi, R. & Marty, B. A determination of the neon isotopic composition of the deep mantle. *Earth Planet. Sci. Lett.* **225**, 77–88 (2004).

28. 28.

French, S. W. & Romanowicz, B. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots. *Nature* **525**, 95–99 (2015).

29. 29.

Ballentine, C. J. & Barfod, D. N. The origin of air-like noble gases in MORB and OIB. *Earth Planet. Sci. Lett.* **180**, 39–48 (2000).

30. 30.

Caffee, M. W. et al. Primordial noble gases from Earth's mantle: identification of primitive volatile component. *Science* **285**, 2115–2118 (1999).

31. 31.

Holland, G. & Ballentine, C. J. Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* **441**, 186–191 (2006).

32. 32.

Bekaert, D. V., Broadley, M. W., Caracausi, A. & Marty, B. Novel insights into the degassing history of Earth's mantle from high precision noble gas analysis of magmatic gas. *Earth Planet. Sci. Lett.* **525**, 115766 (2019).

33. 33.

Parai, R. & Mukhopadhyay, S. The evolution of MORB and plume mantle volatile budgets: constraints from fission Xe isotopes in Southwest Indian Ridge basalts. *Geochem. Geophys. Geosyst.* **16**, 719–735 (2015).

34. 34.

Busemann, H. & Eugster, O. The trapped noble gas component in achondrites. *Meteorit. Planet. Sci.* **37**, 1865–1891 (2002).

35. 35.

Broadley, M. W., Bekaert, D. V., Marty, B., Yamaguchi, A. & Barrat, J.-A. Noble gas variations in ureilites and their implications for ureilite parent body formation. *Geochim. Cosmochim. Acta* **270**, 325–337 (2020).

36. 36.

Pető, M. K., Mukhopadhyay, S. & Kelley, K. A. Heterogeneities from the first 100 million years recorded in deep mantle noble gases from the Northern Lau Back-arc Basin. *Earth Planet. Sci. Lett.* **369–370**, 13–23 (2013).

37. 37.

Ott, U., Begemann, F., Yang, J. & Epstein, S. S-process krypton of variable isotopic composition in the Murchison meteorite. *Nature* **332**, 700–702 (1988).

38. 38.

Lewis, R. S., Amari, S. & Anders, E. Interstellar grains in meteorites: II. SiC and its noble gases. *Geochim. Cosmochim. Acta* **58**, 471–494 (1994).

39. 39.

Dauphas, N. The isotopic nature of the Earth's accreting material through time. *Nature* **541**, 521–524 (2017).

40. 40.

Pepin, R. O. & Porcelli, D. Xenon isotope systematics, giant impacts, and mantle degassing on the early Earth. *Earth Planet. Sci. Lett.* **250**, 470–485 (2006).

41. 41.

Pepin, R. O. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus* **92**, 2–79 (1991).

42. 42.

Carlson, R. W. et al. How did early Earth become our modern world? *Annu. Rev. Earth Planet. Sci.* **42**, 151–178 (2014).

43. 43.

Rubin, M. et al. Krypton isotopes and noble gas abundances in the coma of comet 67P/Churyumov–Gerasimenko. *Sci. Adv.* **4**, eaar6297 (2018).

44. 44.

Mazor, E., Heymann, D. & Anders, E. Noble gases in carbonaceous chondrites. *Geochim. Cosmochim. Acta* **34**, 781–824 (1970).

45. 45.

Heber, V. S. et al. Noble gas composition of the solar wind as collected by the Genesis mission. *Geochim. Cosmochim. Acta* **73**, 7414–7432 (2009).

46. 46.

Pahlevan, K. & Stevenson, D. J. Equilibration in the aftermath of the lunar-forming giant impact. *Earth Planet. Sci. Lett.* **262**, 438–449 (2007).

47. 47.

Lock, S. J. & Stewart, S. T. The structure of terrestrial bodies: Impact heating, corotation limits, and synestias. *J. Geophys. Res. Planets* **122**, 950–982 (2017).

48. 48.

Bekaert, D. V., Broadley, M. W. & Marty, B. The origin and fate of volatile elements on Earth revisited in light of noble gas data obtained from comet 67P/Churyumov–Gerasimenko. *Sci. Rep.* **10**, 5796 (2020).

49. 49.

Ciesla, F. J. & Cuzzi, J. N. The evolution of the water distribution in a viscous protoplanetary disk. *Icarus* **181**, 178–204 (2006).

50. 50.

Raymond, S. N. & Izidoro, A. Origin of water in the inner Solar System: planetesimals scattered inward during Jupiter and Saturn's rapid gas accretion. *Icarus* **297**, 134–148 (2017).

51. 51.

Geist, D. et al. Submarine Fernandina: magmatism at the leading edge of the Galapagos hot spot. *Geochem. Geophys. Geosyst.* **7**, Q12007 (2006).

52. 52.

Peterson, M. E. et al. Submarine basaltic glasses from the Galapagos Archipelago: determining the volatile budget of the mantle plume. *J. Petrol.* **58**, 1419–1450 (2017).

53. 53.

Graham, D. W., Hanan, B. B., Hémond, C., Blichert-Toft, J. & Albarède, F. Helium isotopic textures in Earth's upper mantle. *Geochem. Geophys. Geosyst.* **15**, 2048–2074 (2014).

54. 54.

Colin, A., Moreira, M., Gautheron, C. & Burnard, P. Constraints on the noble gas composition of the deep mantle by bubble-by-bubble analysis of a volcanic glass sample from Iceland. *Chem. Geol.* **417**, 173–183 (2015).

55. 55.

Péron, S., Mukhopadhyay, S. & Huh, M. A new dual stainless steel cryogenic trap for efficient separation of krypton from argon and xenon. *J. Analyt. At. Spectrom.* **35**, 2663–2671 (2020).

56. 56.

Lott, D. E. III Improvements in noble gas separation methodology: a nude cryogenic trap. *Geochem. Geophys. Geosyst.* **2**, 1068 (2001).

57. 57.

Stanley, R. H. R., Baschek, B., Lott, D. E. III & Jenkins, W. J. A new automated method for measuring noble gases and their isotopic ratios in water samples. *Geochem. Geophys. Geosyst.* **10**, Q05008 (2009).

58. 58.

Kunz, J., Staudacher, T. & Allègre, C. J. Plutonium-fission xenon found in Earth's mantle. *Science* **280**, 877–880 (1998).

59. 59.

Moreira, M., Kunz, J. & Allègre, C. J. Rare gas systematics on popping rock: estimates of isotopic and elemental compositions in the upper mantle. *Science* **279**, 1178–1181 (1998).

60. 60.

Parai, R. & Mukhopadhyay, S. Heavy noble gas signatures of the North Atlantic Popping Rock 2IID43: implications for mantle noble gas heterogeneity. *Geochim. Cosmochim. Acta* **294**, 89–105 (2021).

61. 61.

Tucker, J. M., Mukhopadhyay, S. & Schilling, J.-G. The heavy noble gas composition of the depleted MORB mantle (DMM) and its implications for the preservation of heterogeneities in the mantle. *Earth Planet. Sci. Lett.* **355–356**, 244–254 (2012).

62. 62.

Parai, R., Mukhopadhyay, S. & Standish, J. J. Heterogeneous upper mantle Ne, Ar and Xe isotopic compositions and a possible Dupal noble gas signature recorded in basalts from the Southwest Indian Ridge. *Earth Planet. Sci. Lett.* **359–360**, 227–239 (2012).

63. 63.

Moreira, M., Rouchon, V., Muller, E. & Noirez, S. The xenon isotopic signature of the mantle beneath Massif Central. *Geochim. Perspect. Lett.* **6**, 28–32 (2018).

64. 64.

Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry* (Cambridge Univ. Press, 2002).

65. 65.

Huss, G. R. & Lewis, R. S. Presolar diamond, SiC, and graphite in primitive chondrites: abundances as a function of meteorite class and petrologic type. *Geochim. Cosmochim. Acta* **59**, 115–160 (1995).

66. 66.

Pepin, R. O. On noble gas processing in the solar accretion disk. *Space Sci. Rev.* **106**, 211–230 (2003).

67. 67.

Eugster, O., Eberhardt, P. & Geiss, J. Krypton and xenon isotopic composition in three carbonaceous chondrites. *Earth Planet. Sci. Lett.* **3**, 249–257 (1967).

68. 68.

Marti, K. Isotopic composition of trapped krypton and xenon in chondrites. *Earth Planet. Sci. Lett.* **3**, 243–248 (1967).

69. 69.

Manuel, O. K., Wright, R. J., Miller, D. K. & Kuroda, P. K. Heavy noble gases in Leoville: The case for mass fractionated xenon in carbonaceous chondrites. *J. Geophys. Res.* **75**, 5693–5701 (1970).

70. 70.

Krummenacher, D., Merrihue, C. M., Pepin, R. O. & Reynolds, J. H. Meteoritic krypton and barium versus the general isotopic anomalies in meteoritic xenon. *Geochim. Cosmochim. Acta* **26**, 231–249 (1962).

71. 71.

Manuel, O. K., Wright, R. J., Miller, D. K. & Kuroda, P. K. Isotopic compositions of rare gases in the carbonaceous chondrites Mokoia and Allende. *Geochim. Cosmochim. Acta* **36**, 961–983 (1972).

72. 72.

Matsuda, J.-I., Lewis, R. S., Takahashi, H. & Anders, E. Isotopic anomalies of noble gases in meteorites and their origins—VII. C3V carbonaceous chondrites. *Geochim. Cosmochim. Acta* **44**, 1861–1874 (1980).

73. 73.

Basford, J. R., Dragon, J. C., Pepin, R. O., Coscio, M. R. Jr & Murthy, V. R. Krypton and xenon in lunar fines. *Lunar Planet. Sci. Proc.* **4**, 1915–1955 (1973).

74. 74.

Nier, A. O. A redetermination of the relative abundances of the isotopes of neon, krypton, rubidium, xenon and mercury. *Phys. Rev.* **79**, 450–454 (1950).

75. 75.

Nief, G. *Isotopic Abundance Ratios Given for Reference Samples Stocked by the National Bureau of Standards* (ed. Mohler, F.) NBS Technical Note 51 (National Bureau of Standards, 1960).

76. 76.

Eugster, O., Eberhardt, P. & Geiss, J. The isotopic composition of krypton in unequilibrated and gas rich chondrites. *Earth Planet. Sci. Lett.* **2**, 385–393 (1967).

77. 77.

Marti, K., Eberhardt, P. & Geiss, J. Spallation, fission, and neutron capture anomalies in meteoritic krypton and xenon. *Z. Naturforsch. A* **21**, 398–426 (1966).

78. 78.

Eugster, O. Cosmic-ray production rates for  $^3\text{He}$ ,  $^{21}\text{Ne}$ ,  $^{38}\text{Ar}$ ,  $^{83}\text{Kr}$ , and  $^{126}\text{Xe}$  in chondrites based on  $^{81}\text{Kr}$ – $\text{Kr}$  exposure ages. *Geochim. Cosmochim. Acta* **52**, 1649–1662 (1988).

79. 79.

Wieler, R. Cosmic-ray-produced noble gases in meteorites. *Rev. Mineral. Geochem.* **47**, 125–170 (2002).

80. 80.

Eugster, O., Eberhardt, P. & Geiss, J. Isotopic analyses of krypton and xenon in fourteen stone meteorites. *J. Geophys. Res.* **74**, 3874–3896 (1969).

81. 81.

Nakashima, D. & Nakamura, T. Trapped noble gas components and exposure history of the enstatite chondrite ALH84206. *Geochem. J.* **40**, 543–555 (2006).

82. 82.

Okazaki, R., Takaoka, N., Nagao, K. & Nakamura, T. Noble gases in enstatite chondrites released by stepped crushing and heating. *Meteorit. Planet. Sci.* **45**, 339–360 (2010).

83. 83.

Heber, V. S. et al. Isotopic fractionation of solar wind: evidence from fast and slow solar wind collected by the Genesis Mission. *Astrophys. J.* **759**, 121–133 (2012).

84. 84.

Pepin, R. O., Schlutter, D. J., Becker, R. H. & Reisenfeld, D. B. Helium, neon, and argon composition of the solar wind as recorded in gold and other Genesis collector materials. *Geochim. Cosmochim. Acta* **89**, 62–80 (2012).

## Acknowledgements

We thank M. Huh for assistance in the lab; and M. Huyskens, S. Stewart and S. J. Lock for helpful discussions and comments on the manuscript. The He isotope analyses in the OSU noble gas lab were supported by NSF 1763255. The collection of the Fernandina samples and WHOI participation was supported by NSF Ocean Sciences.

## Author information

### Affiliations

1. Department of Earth and Planetary Sciences, University of California, Davis, Davis, CA, USA

Sandrine Péron & Sujoy Mukhopadhyay

2. Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Mark D. Kurz

3. College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA

David W. Graham

## Contributions

S.P. and S.M. designed the study. S.P. carried out the noble gas (Ne, Ar, Kr and Xe) analyses, interpreted the data and wrote the manuscript with feedback from S.M. M.D.K. and D.W.G. provided the samples, discussed the results and contributed to the final manuscript preparation. D.W.G. carried out the He and CO<sub>2</sub> analyses of sample DG2017.

## Corresponding author

Correspondence to [Sandrine Péron](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Hirochika Sumino and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

[Extended Data Fig. 1 Krypton isotopic patterns for each analysis.](#)

Isotopic ratios are in delta notation  $\delta^i\text{Kr} = ((^i\text{Kr}/^{84}\text{Kr})_{\text{sample}}/(^i\text{Kr}/^{84}\text{Kr})_{\text{air}} - 1) \times 10^3$ . **a**, Sample D22B-1. **b**, Sample D22B-2. **c**, Sample D22B-3. **d**, Sample D22A. **e**, Samples DG2017-1 and DG2017-2. Patterns of solar wind<sup>24</sup>, AVCC and Phase Q<sup>25</sup> are shown for reference.

[Source data](#)

### [Extended Data Fig. 2 Krypton–neon isotope plot.](#)

The krypton isotopic compositions of the Galápagos and Iceland mantle sources are not precisely known; our data represent a lower limit as explained in Methods. As such, a range is indicated for the deep mantle spanning the Phase Q and AVCC Kr isotopic compositions. The mixing hyperbolae show that accumulating crushing steps with  $^{20}\text{Ne}/^{22}\text{Ne}$  ratios  $>11.75$  allow to obtain Kr isotopic ratios close to the mantle source ratios.

[Source data](#)

### [Extended Data Fig. 3 Estimation of the probability of the observed deficit in \$^{86}\text{Kr}\$ for the D22 \(Galápagos\) and DG2017 \(Iceland\) samples.](#)

The black rectangle represents the probability (0.01%) that both measured  $^{86}\text{Kr}/^{84}\text{Kr}$  ratios for the Galápagos and Iceland plume sources are higher than the predicted values (Fig. 5). The white rectangle represents the probability that both measured ratios are lower than the predicted values, the blue rectangle that the measured ratio for the Galápagos source is lower than the predicted value with the measured ratio for Iceland being higher than the predicted value, and the green rectangle that the measured ratio for the Iceland source is lower than the predicted value with the measured ratio for Galápagos being higher than the predicted value. There is a 99.9% probability that Earth's deep mantle has a deficit in  $^{86}\text{Kr}$  relative to AVCC.

### [Extended Data Fig. 4 Reproducibility of the air standard for krypton isotopic ratios.](#)

**a**,  $^{78}\text{Kr}/^{84}\text{Kr}$ . **b**,  $^{80}\text{Kr}/^{84}\text{Kr}$ . **c**,  $^{82}\text{Kr}/^{84}\text{Kr}$ . **d**,  $^{83}\text{Kr}/^{84}\text{Kr}$ . **e**,  $^{86}\text{Kr}/^{84}\text{Kr}$ . This set of standards was measured over 13 days and include three sizes of air standard, ranging from  $7.52 \times 10^{-12}$  cc to  $2.37 \times 10^{-11}$  cc of  $^{84}\text{Kr}$ . For the samples, only one size of the air standard was used ( $^{84}\text{Kr}$  of  $2.37 \times 10^{-11}$  cc), measurements of this air standard size show an even better reproducibility (typically of 2‰ for  $^{78}\text{Kr}/^{84}\text{Kr}$ , of 2.5‰ for  $^{80}\text{Kr}/^{84}\text{Kr}$ , of 1‰ for  $^{82}\text{Kr}/^{84}\text{Kr}$ , of 1.5‰ for  $^{83}\text{Kr}/^{84}\text{Kr}$  and of 1‰ for  $^{86}\text{Kr}/^{84}\text{Kr}$ ).

**Extended Data Fig. 5 Three-neon isotope plot for the step-crushing analyses and argon–neon isotope plot for the accumulated gas.**

**a**, Neon isotopic ratios for samples D22B-1, D22B-2, D22B-3 and D22A (Galápagos), compared with literature data<sup>11</sup>. **b**, Neon isotopic ratios for samples DG2017-1 and DG2017-2 (Iceland), compared with literature data<sup>14,15,54</sup>. Supplementary Table 2 indicates for which step heavy noble gases were accumulated. Neon-B<sup>4</sup>, Sun<sup>83</sup>, solar wind<sup>84</sup>. **c**,  $^{40}\text{Ar}/^{36}\text{Ar}$  versus  $^{20}\text{Ne}/^{22}\text{Ne}$  for samples D22B-1, D22B-2, D22B-3 and D22A (Galápagos), compared with literature data<sup>11</sup>. **d**,  $^{40}\text{Ar}/^{36}\text{Ar}$  versus  $^{20}\text{Ne}/^{22}\text{Ne}$  for samples DG2017-1 and DG2017-2 (Iceland) compared with literature data<sup>14,15,54</sup>. The  $^{20}\text{Ne}/^{22}\text{Ne}$  ratios for samples D22B-1, D22B-2, D22B-3, D22A, DG2017-1 and DG2017-2 are the average ratios of the accumulated steps, refer to Supplementary Table 2. The measured  $^{40}\text{Ar}/^{36}\text{Ar}$  ratios as well as the average  $^{20}\text{Ne}/^{22}\text{Ne}$  ratios are consistent with previous measurements for these same samples.

**Source data**

**Extended Data Table 1 Krypton isotopic compositions of samples AHA-NEMO2-D22A and AHA-NEMO2-D22B (hereafter D22A and D22B, Fernandina, Galápagos, respectively) and DG2017 (Midfell, Iceland)**  
**Extended Data Table 2 Xenon abundances and isotopic ratios measured with the accumulation protocol for the Galápagos (AHA-NEMO2-D22A and AHA-NEMO2-D22B) and Iceland (DG2017) samples**  
**Extended Data Table 3 Compilation of carbonaceous, ordinary and enstatite chondrites krypton isotopic data**<sup>67,68,69,70,80,81,82</sup>

**Extended Data Table 4 Argon abundances and isotopic ratios measured with the accumulation protocol for the Galápagos (AHA-NEMO2-D22A and AHA-NEMO2-D22B) and Iceland (DG2017) samples**

**Extended Data Table 5 Results of the accumulation tests with air standard aliquots of  $7.52 \times 10^{-12}$  cc of  $^{84}\text{Kr}$**

## **Supplementary information**

### **Supplementary Tables**

This file contains Supplementary Tables 1 and 2.

### **Source data**

#### **Source Data Fig. 1**

#### **Source Data Fig. 2**

#### **Source Data Fig. 3**

#### **Source Data Fig. 4**

#### **Source Data Fig. 5**

#### **Source Data Extended Data Fig. 1**

#### **Source Data Extended Data Fig. 2**

#### **Source Data Extended Data Fig. 5**

### **Rights and permissions**

#### **Reprints and Permissions**

# About this article

## Cite this article

Péron, S., Mukhopadhyay, S., Kurz, M.D. *et al.* Deep-mantle krypton reveals Earth's early accretion of carbonaceous matter. *Nature* **600**, 462–467 (2021). <https://doi.org/10.1038/s41586-021-04092-z>

- Received: 29 September 2020
- Accepted: 04 October 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04092-z>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04092-z>

- Article
- Open Access
- [Published: 01 December 2021](#)

# Footprint evidence of early hominin locomotor diversity at Laetoli, Tanzania

- [Ellison J. McNutt](#) [ORCID: orcid.org/0000-0002-7723-9103<sup>1,2</sup>](#),
- [Kevin G. Hatala](#) [ORCID: orcid.org/0000-0001-9131-5304<sup>3</sup>](#),
- [Catherine Miller<sup>4,5</sup>](#),
- [James Adams](#) [ORCID: orcid.org/0000-0002-0008-7877<sup>6,7</sup>](#),
- [Jesse Casana](#) [ORCID: orcid.org/0000-0002-4445-1802<sup>4,5</sup>](#),
- [Andrew S. Deane<sup>8</sup>](#),
- [Nathaniel J. Dominy](#) [ORCID: orcid.org/0000-0001-5916-418X<sup>4,5</sup>](#),
- [Kallisti Fabian<sup>9</sup>](#),
- [Luke D. Fannin](#) [ORCID: orcid.org/0000-0003-4810-4442<sup>4,5</sup>](#),
- [Stephen Gaughan<sup>7</sup>](#),
- [Simone V. Gill<sup>10</sup>](#),
- [Josephat Gurtu<sup>9</sup>](#),
- [Ellie Gustafson<sup>11</sup>](#),
- [Austin C. Hill](#) [ORCID: orcid.org/0000-0002-8397-8105<sup>5,12</sup>](#),
- [Camille Johnson<sup>5</sup>](#),
- [Said Kallindo<sup>9</sup>](#),
- [Benjamin Kilham<sup>13</sup>](#),
- [Phoebe Kilham<sup>13</sup>](#),
- [Elizabeth Kim<sup>2</sup>](#),
- [Cynthia Liutkus-Pierce<sup>14</sup>](#),
- [Blaine Maley<sup>15</sup>](#),
- [Anjali Prabhat<sup>5</sup>](#),
- [John Reader<sup>16</sup>](#),
- [Shirley Rubin](#) [ORCID: orcid.org/0000-0002-0520-207X<sup>17</sup>](#),
- [Nathan E. Thompson](#) [ORCID: orcid.org/0000-0002-9273-3636<sup>18</sup>](#),
- [Rebeca Thornburg<sup>11</sup>](#),
- [Erin Marie Williams-Hatala<sup>3</sup>](#),
- [Brian Zimmer<sup>14</sup>](#),

- [Charles M. Musiba](#) ORCID: orcid.org/0000-0002-4297-5178<sup>11,19,20</sup> &
- [Jeremy M. DeSilva](#)<sup>4,5,19</sup>

*Nature* volume 600, pages 468–471 (2021)

- 32k Accesses
- 1 Citations
- 2179 Altmetric
- [Metrics details](#)

## Subjects

- [Archaeology](#)
- [Biological anthropology](#)

## Abstract

Bipedal trackways discovered in 1978 at Laetoli site G, Tanzania and dated to 3.66 million years ago are widely accepted as the oldest unequivocal evidence of obligate bipedalism in the human lineage<sup>1,2,3</sup>. Another trackway discovered two years earlier at nearby site A was partially excavated and attributed to a hominin, but curious affinities with bears (ursids) marginalized its importance to the paleoanthropological community, and the location of these footprints fell into obscurity<sup>3,4,5</sup>. In 2019, we located, excavated and cleaned the site A trackway, producing a digital archive using 3D photogrammetry and laser scanning. Here we compare the footprints at this site with those of American black bears, chimpanzees and humans, and we show that they resemble those of hominins more than ursids. In fact, the narrow step width corroborates the original interpretation of a small, cross-stepping bipedal hominin. However, the inferred foot proportions, gait parameters and 3D morphologies of footprints at site A are readily distinguished from those at site G, indicating that a minimum of two hominin taxa with different feet and gaits coexisted at Laetoli.

[Download PDF](#)

## Main

In 1976, Peter Jones and Philip Leakey discovered five consecutive bipedal footprints at Laetoli site A within locality 7, a 490 m<sup>2</sup> area dated to 3.66 million years ago (Ma)

and featuring 18,400 animal tracks<sup>1,2,3</sup> (Fig. 1). Mary Leakey tentatively suggested that the trackway was made by a hominin<sup>1</sup>. “The footprints,” she wrote, “indicate a rolling and probably slow-moving gait, with the hips swivelling at each step, as opposed to the free-striding gait of modern man [humans].” Leakey and Hay<sup>2</sup> classified the footprints as Hominidae, but with a caveat that “the gait was somewhat shambling, with one foot crossing in front of the other.”

**Fig. 1: Laetoli location and site rediscovery.**

 figure1



**a**, A model of site A generated using photogrammetry showing the five hominin footprints. **b**, Corresponding contour map of the site generated from a 3D surface scan with scale bar. **c**, Map of Laetoli localities 7 and 8, indicating the positions of bipedal trackways A, G and S (redrawn from ref. <sup>49</sup>). **d**, **e**, Topographical maps of the two best preserved A footprints, A2 (**d**) and A3 (**e**).

Unequivocal hominin footprints were discovered at site G two years later, casting doubt on the hominin status of those at site A<sup>3,4,5</sup>. Researchers described the footprints

at site A as “most unusual,”<sup>6</sup> “curiously shaped,”<sup>7</sup> and “enigmatic,”<sup>8</sup> and yet consensus was uniform: they were produced by a plantigrade mammal moving bipedally.

Tuttle<sup>4</sup> advanced three hypotheses to account for the morphology of the footprints and cross-stepping gait (that is, when a foot from each side crosses the midline before touchdown): (1) substrate distortion; (2) they were left by a juvenile bear (ursid); or (3) they are evidence of another hominin species. To test the second possibility, Tuttle<sup>4,9,10,11,12</sup> collected data from circus bears trained to walk bipedally and found that their short steps and relatively wide feet were a close match to the site A footprints, although bipedal bears take wider steps. Furthermore, the fifth digit is typically the largest in ursids, solving the ‘cross-stepping problem’, although Tuttle<sup>4,12</sup> noted that humans do occasionally cross-step. He concluded that “until detailed, naturalistic biometric and kinesiological studies are performed on bipedal bears and barefoot humans, we will have to defer choosing among the hominid and ursid hypotheses on Laetoli individual A”<sup>4</sup>.

Complicating matters further, the internal morphology of the site A footprints was never fully cleaned of matrix infill<sup>1,2,4,5,12</sup>. White and Suwa<sup>8</sup> argued that “reliable identification of these enigmatic prints at Laetoli site A will be impossible until they are more fully cleaned and followed laterally”. Accordingly, we were motivated to relocate and re-excavate site A and conduct a detailed comparative analysis of the prints as well as the locomotion of bears (*Ursus americanus*), chimpanzees (*Pan troglodytes*) and humans to test whether the footprints at site A were left by a hominin or an ursid.

## Rediscovery of site A

Using detailed maps from Leakey and Harris<sup>3</sup>, we identified the proboscidean trail adjacent to the bipedal footprints. We cleared the surrounding overburden until one of us (K.F.) found the A3 footprint. The area was then brushed clean to expose A1–A5, which have experienced no discernible erosion since their initial discovery (Fig. 1, Extended Data Fig. 1). Because the footprint tuff is eroded to the north, we excavated south (87 cm) and east (54 cm) from the heel of A1, but no additional footprints were found ([Supplementary Information](#)).

After brushing sediment from A3, we used a wooden tongue depressor to remove tuff infill left intact during the 1976–1978 field seasons. The hallucial impression is clearly defined and is about 30 mm wide. Crucially, we exposed the impression of the second digit (Extended Data Fig. 2). We removed infill from A2 but could not do so completely without risking damage. Nevertheless, the heel and hallucial impression

are clear. Detailed information from the other footprints (A1, A4 and A5) is limited to estimates of length, width and step length.

Although preservation quality varies within and between A1–A5, there is no evidence that biologically informative metrics were affected by substrate distortion. Adjacent and comingled tracks of other animals (ranging in size from guinea fowl to elephants) show no evidence of distortion to their perimeters or internal morphologies. Given that track surfaces are likely to represent time scales of hours to days<sup>13,14,15</sup>, it is parsimonious to infer similar substrate conditions and taphonomic processes during print formation and subsequent epochs.

## Evaluating ursid and hominin hypotheses

We recorded 50.9 h of video of wild American black bear behaviour. Unsupported bipedal posture and locomotion occurred only 0.09% of the total observation time, of which 59% was postural and 41% was locomotor (Extended Data Fig. 3). In only one instance did a bear take four unassisted bipedal steps. Thus, assuming that our findings are generalizable to other ursids, the probability of observing four consecutive bipedal steps is 0.003%. The low frequency of this behaviour, and the absence of quadrupedal–bipedal transitional footsteps, makes it unlikely, but not impossible, that ursid bipedalism was preserved at site A. Further, Laetoli is devoid of ursid fossils despite the recovery of more than 25,000 fossils attributed to 85 mammalian species<sup>16,17</sup>. If present at all, ursids were rare on the landscape. Although footprint assemblages can include a surprising number of tracks from taxa whose skeletal fossils are rare (for example, relatively high frequencies of bird tracks at Laetoli<sup>3</sup> and at 1.5-Ma sites near Ileret, Kenya<sup>14,18</sup>), there is no clear taphonomic explanation for why ursid tracks would be present but their fossils absent.

In addition, we measured 46 footprints from four bipedally walking wild juvenile black bears specifically chosen because their foot lengths (mean = 145.7 mm) were within 10% of the length of the site A footprints (mean = 161.7 mm). Additionally, we measured the footprints of chimpanzees produced during quadrupedalism ( $n = 54$  from 46 adults; Ngamba Island Chimpanzee Sanctuary, Uganda) and during bipedalism ( $n = 44$  from two subadults; Stony Brook University, USA). We compared these data with human barefoot footprints produced under three conditions: (1) habitually shod ( $n = 654$ ) walking on a plantar pressure mat<sup>19</sup>; (2) habitually unshod or minimally shod ( $n = 41$ ) walking in deformable mud<sup>20,21</sup>; and (3) Late Pleistocene tracks ( $n = 113$ ) from Engare Sero, Tanzania, formed in reworked volcanic ash<sup>22,23</sup> (summary in Extended Data Table 1).

We concur with others<sup>4,12</sup> that the ratios of footprint dimensions (for example, heel and forefoot width) to step length observed at site A fall within the ursid range (Extended Data Fig. 4a, b). Yet, for these same measures, site A is also chimpanzee-like and moderately similar to definitive hominin footprints from sites G and S. It follows that the site A individual was taking short steps—as occurs when humans walk slowly or over a slippery substrate<sup>24</sup>—not that the gait was ursid-like.

With additional infill removed from A2 and A3, the perimeter dimensions are decidedly hominin-like with wide heel impressions relative to forefoot width (Extended Data Fig. 4c). By contrast, chimpanzees and bears have relatively narrow heels. Furthermore, with the tracks fully excavated and cleaned, we found no evidence for claw impressions, although they are sometimes absent from ursid footprints<sup>12</sup>. Here, impressions were absent from 31% of ursid footprints. To test whether A3 was produced by a hominin left foot or an ursid right foot, we compared the width of the hallux to the second digit in human ( $n = 30$ ) and chimpanzee ( $n = 50$ ) footprints, and the fifth digit to the fourth in bear ( $n = 5$ ) tracks. The A3 toe impressions match the distinctive proportions of humans and chimpanzees rather than those of bears (Extended Data Fig. 5).

By establishing that A3 is a left hominin foot, we can now confirm that cross-stepping occurred. Cross-stepping was never observed in our comparative sample, but humans do it occasionally<sup>25,26</sup> as a compensatory strategy for re-establishing balance after a perturbation<sup>27</sup>. In fact, we suggest that cross-stepping supports the hypothesis that the site A footprints were left by a hominin. Cross-stepping is improbable, and perhaps impossible, when bears or chimpanzees walk bipedally. They produce large mediolateral excursions of their centre of mass<sup>28</sup> and walk with highly abducted hips<sup>29</sup>, resulting in a high ratio between stride width and step length. Conversely, human cross-stepping is enabled by their reduced mediolateral centre of mass and body motions, adducted hips and bicondylar angle (that is, valgus knees), resulting in a low corresponding ratio, as expressed in every trackway at Laetoli.

The relative step widths of footprints from sites G and S fall squarely within the modern human distribution (Extended Data Figs. 4d, 5). The site A footprints lie outside the distributions of humans, chimpanzees and bears but are most like humans. This result indicates that the maker of the site A footprints had either valgus knees or adducted hips, or both. The presence of either bipedal trait argues for a hominin maker of the site A footprints.

## Which hominin?

It is generally accepted that *Australopithecus afarensis* produced the footprints at sites G and S<sup>8</sup> (but see refs. [12,30,31,32](#)). It is thus tempting to assign the site A tracks to *A. afarensis*; however, this premise requires an examination of foot ontogeny and intraspecific morphological variation that takes into account the mounting fossil evidence of locomotor (and presumably taxonomic) diversity among Pliocene hominins (for example, in ref. [33](#)).

Standing between 101 and 104 cm tall (from equations in Dingwall et al.<sup>34</sup>), the maker of the site A footprints was smaller in height than other Laetoli trackmakers, which ranged from 111–116 cm (site G1) to 161–168 cm (site S1)<sup>35</sup>. It is plausible that the site A tracks were made by a juvenile *A. afarensis*, but this hypothesis is undermined by a distinct footprint morphology from those at sites G or S.

The ratio of foot width and length follows a different ontogenetic trajectory in humans and chimpanzees; human feet are consistently narrower than chimpanzee feet (Fig. [2](#)). Footprints made by unshod humans from both modern times and the Pleistocene are slightly wider than those made by shod humans in industrialized populations. The undistorted footprints from sites G and S fall within the human distribution. A3 is more chimpanzee-like in being wide compared with its length (Fig. [2a](#), Extended Data Fig. [2](#)). In chimpanzees, this wider footprint shape is, in part, driven by the greater divergence of the hallux. We thus measured hallucial divergence as a ratio of the distance between the centre of the impression made by the first and second digits and the length of the footprint. For this metric, humans and chimpanzees are clearly distinct. The best-defined site G footprints overlap the human distribution, whereas the A3 footprint does not—it possesses a slightly more divergent hallux than humans and site G, although not nearly as divergent as those of chimpanzees (Extended Data Fig. [6](#)). This finding alone does not rule out a juvenile *A. afarensis*, given the foot from Dikika (Afar, Ethiopia) has a slightly more divergent and mobile hallux than its adult counterparts<sup>36</sup>.

**Fig. 2: Evidence against Laetoli A belonging to *A. afarensis*.**

---

 **figure2**

**a**, Images of Laetoli A3 (top) length matched to Laetoli G1 (bottom). G1 print is reproduced with permission from Science Photo Library. **b**, Plot comparing foot length to forefoot width in adult and juvenile humans and bipedal chimpanzees, the Pleistocene human footprints at Engare Sero, and the Laetoli trackways. Lines represent ordinary least-squares regression and shaded bands represent 95% confidence interval of the regression. Data were analysed from the total number of individual footprints indicated in the figure legend. Laetoli footprint markers are slightly enlarged for clarity. Data sources match Extended Data Table 1. **c**, Histogram of Mahalanobis distances between the mean modern human footprint and the averages of two randomly drawn human footprints (grey;  $n = 245$ , resampled 1,000 times) and two randomly drawn bipedal chimpanzee footprints (yellow;  $n = 45$ , resampled 1,000

times). The blue and orange lines represent the distances of mean Laetoli G1 ( $n = 5$ , with 10 unique two-track combinations) and S1 ( $n = 2$ ) two-track samples, respectively. All Laetoli G1 and S1 samples fall within the human distribution. The green line indicates the distance between the mean Laetoli A track ( $n = 2$ ) and the human mean, falling squarely within the chimpanzee distribution.

### Source data.

To explore other instructive traits, we compared proportional toe depth ratios as described by Raichlen and Gordon<sup>37</sup> ([Supplementary Methods](#)). The mean value for site A ( $-0.191$ ) is distinct from Laetoli G1 and humans using a bent-hip bent-knee gait but overlaps the low end of variation in Laetoli S. Additionally, tracks A1–A3 evince a raised ridge of hardened ash between the heel and lateral forefoot. It is unclear whether this ridge is evidence of substrate shearing<sup>38</sup> or midfoot mobility—a characteristic absent from the Laetoli G and S footprints<sup>39</sup> (but see ref. <sup>40</sup>) and inconsistent with *A. afarensis* pedal remains<sup>41</sup>.

Finally, we tested whether we could randomly sample footprints with internal topography similar to the various Laetoli footprints from those of humans or chimpanzees (following Hatala et al.<sup>42</sup>). Fig. 2 illustrates how the tracks from sites G and S can be encompassed in the range of resampled unshod human footprint variation, whereas the average morphology of the A2 and A3 tracks is distinct from the footprints of habitually unshod humans and those at sites G and S. In fact, they fit comfortably within the resampled chimpanzee distribution, being as distinct as chimpanzee tracks are from the morphologies of barefoot human tracks. One possible explanation for such different footprint morphology is that the site A footprints were made by a cross-stepping *A. afarensis*. We tested this hypothesis by comparing the footprints of humans ( $n = 10$ ) walking with their preferred gait and then cross-stepping. We found that normal and cross-stepping human footprints differ minimally and do not match in magnitude or direction the differences between the site G and S prints and the site A prints (Extended Data Fig. 7, [Supplementary Information](#)).

We therefore conclude that the site A footprints were made by a bipedal hominin with a distinct and presumably more primitive foot than *A. afarensis*. The gross shape of the foot is chimpanzee-like, with slight hallucial divergence and perhaps some midfoot mobility. However, the site A individual was walking bipedally with a narrow step width indicative of either a valgus knee, adducted hips, or both. This combination of foot morphology and gait kinematics inferred from the preserved footprints precludes them from having been made by *A. afarensis*.

Evidence is building for taxonomic diversity in hominins during the Pliocene<sup>43,44</sup>, including at Laetoli<sup>45,46,47</sup>, but these hominins did not walk with morphologically

identical feet<sup>48</sup>. For example, the BRT-VP-2/73 foot from the 3.4 Ma site of Woranso-Mille, Ethiopia demonstrates that at least two different foot morphs co-existed in the Afar Depression during the Pliocene<sup>33</sup>. We suggest that footprint evidence for hominin locomotor diversity is similarly present at Laetoli, Tanzania—and has been since the discovery of the site A trackway in the 1970s.

## Methods

### Wild bear behavioural data

Wild black bear behaviour was quantified using video data recorded by B.K. over the course of several years at his ongoing field site in Lyme, New Hampshire, USA. Video data captured bears of different ages (cubs, adolescents and adults). Bears were present on screen for a total of 50 h 55 min 18 s. For each terrestrial bipedal incident, the length of the event, the approximate age of the bear and the number of steps were recorded. Additionally, steps were evaluated on whether they were completed independently, or the individuals used other environmental objects for balance.

### Comparative kinematic data

Comparative kinematic data were collected on three species: *U. americanus*, *P. troglodytes*, and *Homo sapiens*. For bears and chimpanzees, the sample size included all available individuals housed at each location. For the human sample size information see below. Randomization was not relevant to our study as we were interested in measuring footprint characteristics from whole sample populations, as opposed to comparisons within those populations. Blinding was not relevant to the data collected on the non-human comparative species (for example, bears and chimpanzees) nor to the data collection on fossilized footprints. The human participants were unaware of the site A tracks at Laetoli and therefore had no knowledge of how the data obtained from their footprints would be used in this study.

#### *Ursus americanus*

Data were collected on four juvenile semi-wild *U. americanus* ( $n = 3$  male, 1 female), whose feet were within 10% of the length (average foot length = 145.7 mm) of the recorded footprints of Laetoli site A (average foot length = 161.7 mm). These orphaned, approximately 20-kg bears were located at the Kilham Bear Center (Lyme, NH), awaiting reintroduction to the wild. This study examined the bears between the ages of 5–8 months old. Our protocol was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of Dartmouth College. The bears were enticed to independently walk bipedally through a constructed mud trackway for either

an applesauce or maple syrup reward (Extended Data Fig. 3). Measurements were collected on the footprints, including foot length, heel width, forefoot width, step length and stride width using the definitions from Tuttle<sup>4</sup>. For a subset of footprints ( $n = 5$ ), the width of the impression for the 1st, 2nd, 4th and 5th digits were measured. The presence or absence of claw impressions was also documented.

### ***Pan troglodytes***

Data for extant chimpanzees were extracted from three sources to collect all the relevant gait metrics. Two published datasets examined the same two subadult individuals housed at Stony Brook University. The third set were recorded on semi-wild individuals ( $n = 46$ ), using a plantar pressure mat at the Ngamba Island Chimpanzee Sanctuary (Entebbe, Uganda). While this third data set increases sample size and captures intraspecific variation, we recognize that plantar pressure data do not always align perfectly with footprints made in a deformable substrate<sup>20</sup>.

### **Stride width data and step length comparisons**

Chimpanzee stride width data were taken from Thompson et al.<sup>28</sup> on two subadult male chimpanzees ( $7.0 \pm 0.1$  years of age;  $34.8 \pm 1.2$  kg) and were supplemented with step length data for the same steps. Three-dimensional kinematic methods and step width calculation have been described previously<sup>28</sup>. Step length was calculated as the distance between left and right calcaneus markers in the sagittal plane during consecutive hind limb midstance periods. Chimpanzee step lengths are typically asymmetric, so step length was averaged over the two consecutive steps which defined the stride.

### **Forefoot width, heel width and foot length comparisons**

Footprint dimensions and stride length data were recorded on the same two subadult male chimpanzees as above, though at a slightly younger age (6.5 and 6.9 years of age, 30.7 and 27.8 kg, respectively). The experimental design is described in detail elsewhere<sup>42</sup>. In brief, chimpanzees traversed a runway, at the centre of which was a pressure mat (RSScan International) and a container of hydrated sediment in which the chimpanzees could produce footprints. This sediment was taken directly from a layer that preserves 1.5 Ma hominin footprints near Ileret, Kenya<sup>50</sup>. Laterally positioned video cameras were used to record the chimpanzees as they walked along this trackway and produced footprints. Two digitization softwares, MaxTRAQ Lite+ (v. 2.4.0.3) (Innovisions Systems) and ImageJ v.1.47<sup>51</sup>, were used to quantify various aspects of their gaits, including stride length. Tape measures and digital callipers were used to directly measure the external dimensions of each chimpanzee's feet. Scaled

photographs were taken of the footprints produced in each trial, and these were later measured using ImageJ software.

### **Forefoot width, width of digits 1 and 2, divergence ratio, and foot length comparisons**

Data were collected by E.J.M. at the Ngamba Island Chimpanzee sanctuary (Entebbe, Uganda) managed by the Chimpanzee Sanctuary and Wildlife Conservation Trust (CSWCT) using procedures approved by the Dartmouth College IACUC. The Tekscan plantar pressure mat (PPM) was positioned within a walkway connecting the overnight enclosure to the open forest habitat, near a gate and underneath a solid cross section to help prevent individuals from jumping over the mat using the ceiling bars. This location was determined using the expertise of the sanctuary keepers. The animals were first introduced to a mat shell that was lacking the internal sensors, to habituate them to the novel stimulus. All subsequent data were collected using both the empty and real PPMs, positioned such that they covered the entire width of the walkway to force individuals to walk across one of the two mats. Both mats were covered with thin green sacks to help disguise them from the chimpanzees and facilitate faster removal if necessary. It was determined that the southeast facing direction was the preferred path for the chimpanzees and the sensor-containing PPM was always positioned there from the second collection onwards. Data were collected twice a day; once in the morning (between 06:45 and 08:00) when the chimpanzees were headed to the forest for their first feeding, and once at 18:00, when the chimpanzees were headed into the overnight enclosure to sleep and receive their last feeding. Data were collected on 46 adult chimpanzees (18 male, 28 female, ages 12–36 years). A subset of 54 dynamic pressure records was analysed. Using the associated Tekscan PPM software, Footmat Research (v. 7.10), the pressure recordings were analysed to determine foot length, forefoot width, the width digits 1 and 2, and the linear distance between the centre of digits 1 and 2. The divergence ratio was calculated by dividing the distance between digits 1 and 2 by the individual's foot length.

### ***Homo sapiens***

Data were extracted from previous studies of two modern human populations in order to collect all the relevant foot, footprint and gait metrics.

### **Stride width and step length comparisons**

Data were taken on 654 participants, recruited through the Living Laboratory at the Boston Museum of Science<sup>19</sup>. Sample size was determined by museum visitor traffic and willingness to participate in a scientific study. In brief, this dataset included 73

children between the ages of 2 and 7 years old (29 female and 44 male) and 581 individuals (366 female and 215 male) between the ages of 8 and 80 years. A pressure-sensitive gait carpet (6.1 m long × 0.89 m wide) with a spatial resolution of 1.27 cm and collecting data at 120 Hz (GAITRite) was used to collect stride length and stride width. For a subset of 33 adults, additional data were collected using a Tekscan PPM and analysed with FootMat Research (v. 7.10) to calculate foot length, the width of digits 1 and 2, and the linear distance between the centre of digits 1 and 2. These measurements were used to calculate a divergence ratio as described above in '*P. troglodytes*'.

### **Forefoot width, heel width and foot length comparisons**

Footprint dimensions and stride length data for 29 Daasanach adults (15 male and 14 female, ages 18–47) and 12 children (10 male and 2 female, ages 4–15), who live near the town of Ileret, Kenya and grew up either habitually unshod or minimally shod were taken from Hatala et al.<sup>20,21</sup>. Details of the experimental protocol largely mirrored the procedures described above. In brief, subjects generated footprints while walking through a rehydrated sample of the same sediments that preserve 1.5 Ma hominin tracks near Ileret. Video cameras were used to record subjects as they produced footprints, and two digitization software packages (MaxTRAQ Lite+ v. 2.4.0.3 and ImageJ v.1.47) were used to measure stride lengths and other kinematic variables. The external dimensions of subjects' feet were directly measured with tape measures and digital callipers. Scaled photographs of the footprints produced in each trial were measured using ImageJ.

### **Human cross-stepping footprint experiments**

Experiments were carried out by K.G.H. and E.M.W.-H. to investigate whether and how cross-stepping kinematics influence the perimeter dimensions and internal topologies of an individual's footprints. We could thereby evaluate whether the size and shape of the Laetoli site A tracks could have been generated by a hominin with feet similar to those who left tracks at sites G and S, but while cross-stepping. Detailed methods are provided in [Supplementary Information](#). In brief, ten adult subjects (including six female, three male, and one non-binary between 19 and 52 years old) each completed ten trials in which they produced tracks in sedimentary conditions meant to mimic those at Laetoli<sup>37,52</sup>. Sample size was determined by availability and willingness to participate in the study. Five trials were completed with a normal, self-selected walking gait and another five were completed with a cross-stepping gait, as inferred for the Laetoli site A trackmaker. In each trial, a focal footprint was selected, measured in situ, and photographed (25–30 photos per footprint). The lengths and widths of the steps bracketing the track were also measured. Photographs were used to

generate 3D models of the tracks using Agisoft Metashape software (v.1.7.3), and average normal and cross-stepping tracks were generated for each subject using DigTrace Pro (v.1.8.1)<sup>53</sup>. Lengths and widths of these averaged tracks were measured using Geomagic Wrap (v. 2021.0.0) (3D Systems). Regional depths were measured and evaluated using the same methods described below ('Comparative analyses of Laetoli footprint shapes'). Within-subject comparisons enabled us to understand how cross-stepping influenced the dimensions of the perimeter and the internal topology of a subject's footprints.

## Fossil footprint data and analysis

Comparative metrics were quantified from a set of modern human footprints from the Late Pleistocene at Engare Sero, Tanzania. These footprints are an important comparison with the Laetoli footprints, as they were generated in a similar circumstance (footprints in volcanic ash) and represent an early population of unshod modern humans.

A scaled 3D orthophoto of the Engare Sero site was created via photogrammetry by B.Z. and C.L.-P. to visualize the distribution of footprint trackways across the entire site using Agisoft Photoscan (now Agisoft Metashape v. 1.4.4). The model was created from hundreds of photos originally taken by the Smithsonian 3D Digitization Program in 2010. Measurements that were defined in Tuttle<sup>4</sup>, were taken from the fossil tracks using the software, ImageJ (v. 1.49), and included foot length, forefoot width, heel width and step width of each footprint at Engare Sero. In some cases, partial footprints were included for measurement, as long as they included the requisite landmarks for those measurements. Overall, data were collected from 151 footprints at the Engare Sero site. Of the 151 footprints, 61 footprints were considered partial footprints and 90 footprints were considered complete footprints. From these, 67 step length and stride width measurements and 105 heel width and ball width measurements were included in our analyses.

All measurements for Laetoli trackways G and S were obtained from published sources<sup>4,32,42</sup>. Box and whisker plots and bivariate graphs (using ggplot2<sup>54</sup>) were generated using R (v. 3.6.1), while the table and pie chart were generated using Microsoft Excel (v. 2102).

## Comparative analyses of Laetoli footprint shapes

Comparative analyses followed methods similar to resampling analyses published previously<sup>42</sup>. In brief, the human comparative sample included 245 footprints produced by 29 adult and 12 juvenile habitually unshod Daasanach individuals traveling at walking speeds. The chimpanzee comparative sample included 45

footprints produced by two individuals walking bipedally. Laetoli samples included only the best-preserved tracks from each site, leaving samples of five footprints from site G that were described by their original excavators as free from taphonomic damage that would obscure track topology (G1-25, G1-27, G1-33, G1-34 and G1-35), and two from site S (L8-S1-2 and L8-S1-4). For site A, we included tracks A2 and A3, as these were the only two for which we were relatively confident in identifying regions of interest across the entire track. Larger sample sizes would be desirable, but we did not want to sacrifice data quality for quantity by including tracks that were overprinted or that did not appear to represent complete foot anatomy. We did not rely on parametric statistical tests for which larger sample sizes would be a necessity, and instead used an analytical approach that could handle smaller sets of observations (see below).

For each experimental and fossil footprint, 3D models were constructed using photogrammetry, through a variety of methods described here for Laetoli site A and elsewhere by the authors for other samples<sup>20,21,35,42</sup>. Using Geomagic Wrap (v. 2021.0.0) (3D Systems), a best-fit plane was fit to the undisturbed substrate surrounding each track, and this was fixed to the  $xy$  plane in world coordinate space. In this orientation, depths of the footprint were measured in the regions of the medial and lateral heel, medial and lateral midfoot, and all five metatarsal heads and toes. Raw depth measurements were normalized, within each footprint, to a scale of 0 to 1 in order to compare the topologies of footprints that may vary in depth. However, a Wilcoxon signed-rank test showed that, overall, human and Laetoli track samples did not differ significantly in their depths ( $P = 0.08$ ). Within-subject means of the 14 normalized depth measurements were calculated, and a between-subject covariance matrix was created using the subject averages for normalized depths at each of the 14 measured regions. An overall ‘human mean footprint’ was also computed by averaging the within-subject mean normalized depths, and this represented a measure of central tendency as described below.

To represent the range of observed variation in human footprint topography, for 1,000 iterations we randomly sampled a human subject and drew a sample of two of their footprints. We then averaged the normalized depths of those two footprints and computed the Mahalanobis distance (using the between-subject covariance matrix) between this track and the mean of all other subjects’ footprints. Also, for 1,000 iterations we selected a random chimpanzee subject, drew a random sample of two of their footprints and computed the Mahalanobis distance between the average of those tracks and the overall mean human footprint. For the Laetoli tracks, site A and site S samples only included two tracks, so these were simply averaged and the Mahalanobis distance was calculated between each averaged track and the mean human footprint. For Laetoli site G, all possible two-track combinations (ten) were drawn from the sample described above, and the Mahalanobis distance was calculated between the

averaged track from each combination and the human mean. In all cases, we calculated multivariate distances using the human between-subject covariance matrix (that is, treating the chimpanzee and fossil tracks as if they came from different human subjects). All analyses described above, and the histogram displaying multivariate distances (Fig. 2), were generated using R (v. 3.6.1), with custom scripts and functions from the dplyr<sup>55</sup>, ggplot2<sup>54</sup> and reshape2<sup>56</sup> packages.

## Photogrammetry

While casts of the site A bipedal footprints existed at one point, all our attempts to locate them (see Acknowledgements for a complete list) were unsuccessful. Prior to our fieldwork at Laetoli in 2019, we modelled the original trackway photogrammetrically using extant photography from the site. Original photography of Laetoli site A was taken by J.R. We obtained his photographs of trackway A through Science Photo Library. All photographs were taken with a Nikon F2 on 35-mm Kodachrome slide film. Digital scans from these slides were used to produce a 3D model of the Laetoli A footprints. Unfortunately, since the images were taken in 1977, they were not recorded with modern photogrammetry processing in mind. Several features of the digitized images limit successful and accurate construction of a 3D model. First, there are only four images of the footprints. One of these images has noticeably different exposure settings that caused significant alignment problems during processing, and thus was excluded. All images were shot at oblique angles, from a relatively narrow range of camera positions. A yellow string defining the site grid lies over one of the footprints, obscuring part of it, and casting a shadow. The images were all taken relatively early in the day, so there are shadows within each footprint that create strong contrasts. The slides were digitized at 4,000 dpi, but they were not scanned with specialized equipment to guarantee geometric accuracy, and this potentially introduced more sources of distortion.

However, despite the limitations of the images, it was possible to extract 3D data for the Laetoli A footprints. All processing was done in Agisoft Photoscan Pro (v. 1.7.1). The standard processing steps (align photos, build dense cloud, build mesh, build texture) were run to produce a 3D model, though the process had to be done iteratively to remove noise, ensure accurate alignment of the photos, scale the model appropriately using published measurements, add manual tie points, and refine the model. A DEM (digital elevation model) and orthophotograph were exported for further visualization and analysis in ArcGIS (v. 10.6.1). The 3D model was also exported to Autodesk Meshmixer (v. 3.5.474) to create a ‘watertight’ 3D volume that could be 3D printed for further visualization (1977 model is hosted on Morphosource, ID: 000390119). Photogrammetric reconstruction was validated using published measurements of the footprints. It is important to note however, that there were no published measurements for the depths of the footprints and that the internal anatomy

of this reconstruction is potentially misleading because of the incomplete excavation of the footprints<sup>[2,8,12](#)</sup>.

A second, more accurate reconstruction was done using photogrammetry from the re-excavated site A bipedal trackway using 57 images taken in June 2019. The images were captured in a systematic manner using a Nikon D7000 camera and Nikon DX AF-S Nikkor 18–105 mm lens. All photos were taken by hand, from an eye level, while walking a series of transects, across the area of interest. Spacing between shots was kept low to ensure a minimum of approximately 65% overlap between adjacent images. All processing was done using Agisoft Photoscan Pro/Metashape Pro (v. 1.7.1). Standard processing steps (for example, as described<sup>[15,51](#)</sup>) were taken to create a 3D model of the A trail. This included photo alignment, manual editing of the sparse cloud to remove points with high ‘reprojection uncertainty’, building a dense cloud, building a mesh, refining the mesh, then building a texture. During processing, images were checked for sharpness using the ‘image quality’ tool and any images with significantly lower quality were removed. The model was scaled to the real-world using scale bars placed across the region of interest. Finally, an orthophotograph as well as a DEM (digital elevation model) were exported as geotiffs into ArcGIS in an arbitrary local coordinate system for further analysis (2019 model hosted on Morphosource, ID: 000390114).

To generate contour maps, two approaches were used. First, starting with the raw stereolithography scans (.stl file format), Ultimaker Cura software (v.4.8.0) was used to rotate the raw scans and align them with x and y axes. This was a manual process. These rotations were exported to binary-format .stl files. The rotated files were then run through an R script using R version 4.0.3. The R script uses the tidyverse and rgl libraries to load the .stl files into R-friendly dataframes and plot them as contours using ggplot’s geom\_contour function. The script is available through GitHub.

Using a second approach, the .stl files were brought into Cloud Compare (v. 2.11.3) to check model orientation. If necessary, models were reoriented to allow the local ground surface to be level using the "level" tool, and then the files were exported. The correctly oriented model was imported into SAGA GIS using the import stereo lithograph file (STL) tool. This tool converts the .stl directly to a DEM raster. The rasters were checked in SAGA and a hillshade generated with the analytic hillshading function using the standard sun position setting of 315° azimuth and 45° height. Both the DEM and hillshade were then exported as geotiffs. These geotiffs were imported into ArcGIS for visualization. The DEM was colored using a red-blue colour ramp to indicate relative depth and this was layered onto the hillshade raster using the NAGI fusion method<sup>[57](#)</sup> (Extended Data Fig. [8](#)). Cloud compare was used to quantify erosive alterations to the site A footprints from 1977 to 2019.

## 3-D surface scanning

Three-dimensional surface scans of Laetoli A and plaster casts of bear prints were collected using a Creaform Go! Scan 50.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Previously published data were obtained from refs. [4,19,20,21,28,32,42](#). All other data supporting the findings of this study are available within the paper and its [supplementary information files](#). The photogrammetric reconstruction of the Laetoli A trackway based on three original photographs from the 1977 expedition is available on Morphosource (<https://www.morphosource.org>; ID: 000390119). The photogrammetric reconstruction of Laetoli A trackway using 57 photographs taken of the re-excavated Laetoli site A footprints in 2019 is available on Morphosource (ID: 000390114). [Source data](#) are provided with this paper.

## References

1. 1.  
Leakey, M. Pliocene footprints at Laetoli, northern Tanzania. *Antiquity* **52**, 133 (1978).
2. 2.  
Leakey, M. D. & Hay, R. L. Pliocene footprints in the Laetoli Beds at Laetoli, northern Tanzania. *Nature* **278**, 317–323 (1979).
3. 3.  
Leakey, M. D. & Harris, J. M. (eds) *Laetoli, a Pliocene Site in Northern Tanzania* (Clarendon, 1987).
4. 4.  
Tuttle, R. H. in *Laetoli: A Pliocene Site in Northern Tanzania* (eds Leakey, M. D. & Harris, J. M.) 503–523 (Clarendon, 1987).

5. 5.

Day, M. H. in *Origine(s) de la Bipedie Chez les Hominides* (eds Coppens Y. & Senut B.) 199–213 (Editions du CNRS, 1991).

6. 6.

Leakey, M. *Disclosing the Past: An Autobiography* (Doubleday, 1984).

7. 7.

Leakey, M. D. Footprints in the ashes of time. *Nat. Geo.* **155**, 446–457 (1979).

8. 8.

White, T. D. & Suwa, G. Hominid footprints at Laetoli: facts and interpretations. *Am. J. Phys. Anthropol.* **72**, 485–514 (1987).

9. 9.

Tuttle, R. H. Bear facts and Laetoli impressions. *Am. J. Phys. Anthropol.* **63**, 230–230 (1984).

10. 10.

Tuttle, R. H. in *Hominid Evolution: Past, Present, and Future* 129–133 (Alan R. Liss, 1985).

11. 11.

Tuttle, R. H. The pitted pattern of Laetoli feet. *J. Nat. Hist.* **99**, 60–65 (1990).

12. 12.

Tuttle, R., Webb, D. & Tuttle, N. in *Origine(s) de la Bipedie Chez les Hominides* (eds Coppens Y. & Senut B.) 187–198 (Editions du CNRS, 1991).

13. 13.

Cohen, A., Lockley, M., Halfpenny, J. & Michel, A. E. Modern vertebrate track taphonomy at Lake Manyara, Tanzania. *Palaeos* **6**, 371–389 (1991).

14. 14.

Roach, N. T. et al. Pleistocene footprints show intensive use of lake margin habitats by *Homo erectus* groups. *Sci Rep.* **6**, 26374 (2016).

15. 15.

Zimmer, B. et al. Using differential structure-from-motion photogrammetry to quantify erosion at the Engare Sero footprint site, Tanzania. *Quat. Sci. Rev.* **198**, 226–241 (2018).

16. 16.

Su, D. F. & Harrison, T. Ecological implications of the relative rarity of fossil hominins at Laetoli. *J. Hum. Evol.* **55**, 672–681 (2008).

17. 17.

Harrison, T. in *Paleontology and Geology of Laetoli: Human Evolution in Context* 1–14 (Springer, 2011).

18. 18.

Roach, N. T. et al. Pleistocene animal communities of a 1.5 million-year-old lake margin grassland and their relationship to *Homo erectus* paleoecology. *J. Hum. Evol.* **122**, 70–83 (2018).

19. 19.

DeSilva, J. M. & Gill, S. V. Brief communication: a midtarsal (midfoot) break in the human foot. *Am. J. Phys. Anthropol.* **151**, 495–499 (2013).

20. 20.

Hatala, K. G., Dingwall, H. L., Wunderlich, R. E. & Richmond, B. G. The relationship between plantar pressure and footprint shape. *J. Hum. Evol.* **65**, 21–28 (2013).

21. 21.

Hatala, K. G., Wunderlich, R. E., Dingwall, H. L. & Richmond, B. G. Interpreting locomotor biomechanics from the morphology of human footprints. *J. Hum. Evol.* **90**, 38–48 (2016).

22. 22.

Liutkus-Pierce, C. M. et al. Radioisotopic age, formation, and preservation of Late Pleistocene human footprints at Engare Sero, Tanzania. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **463**, 68–82 (2016).

23. 23.

Hatala, K. G. et al. Snapshots of human anatomy, locomotion, and behavior from Late Pleistocene footprints at Engare Sero, Tanzania. *Sci. Rep.* **10**, 7740 (2020).

24. 24.

Cham, R. & Redfern, M. S. Changes in gait when anticipating slippery floors. *Gait Posture* **15**, 159–171 (2002).

25. 25.

Fukushima, T. Angle and distance between steps. *J. Bone. Jt. Surg.* **37**, 1311 (1955).

26. 26.

Murray, M. P., Drought, A. B. & Kory, R. C. Walking patterns of normal men. *J. Bone Jt. Surg.* **46**, 335–360 (1964).

27. 27.

Maki, B. E. & McIlroy, W. E. Control of rapid limb movements for balance recovery: age-related changes and implications for fall prevention. *Age Ageing* **35**, ii12–ii18 (2006).

28. 28.

Thompson, N. E., O'Neill, M. C., Holowka, N. B. & Demes, B. Step width and frontal plane trunk motion in bipedal chimpanzee and human walking. *J. Hum. Evol.* **125**, 27–37 (2018).

29. 29.

O'Neill, M. C., Demes, B., Thompson, N. E. & Umberger, B. R. Three-dimensional kinematics and the origin of the hominin walking stride. *J. R. Soc. Interface* **15**, 20180205 (2018).

30. 30.

Tuttle, R. H. Evolution of hominid bipedalism and prehensile capabilities. *Philos. Trans. R. Soc. Lond. B.* **292**, 89–94 (1981).

31. 31.

Harcourt-Smith, W. E. Did *Australopithecus afarensis* make the Laetoli footprint trail?—New insights into an old problem. *Am. J. Phys. Anthropol.* **S40**, 116 (2005).

32. 32.

Pelissero, A. J. *A Comparative Analysis of Newly Discovered Pliocene Hominin Footprints from Laetoli, Tanzania* (Univ. Colorado at Denver, 2017).

33. 33.

Haile-Selassie, Y. et al. A new hominin foot from Ethiopia shows multiple Pliocene bipedal adaptations. *Nature* **483**, 565–569 (2012).

34. 34.

Dingwall, H. L., Hatala, K. G., Wunderlich, R. E. & Richmond, B. G. Hominin stature, body mass, and walking speed estimates based on 1.5 million-year-old fossil footprints at Ileret, Kenya. *J. Hum. Evol.* **64**, 556–568 (2013).

35. 35.

Masao, F. T. et al. New footprints from Laetoli (Tanzania) provide evidence for marked body size variation in early hominins. *eLife* **5**, e19568 (2016).

36. 36.

DeSilva, J. M., Gill, C. M., Prang, T. C., Bredella, M. A. & Alemseged, Z. A nearly complete foot from Dikika, Ethiopia and its implications for the ontogeny and function of *Australopithecus afarensis*. *Sci. Adv.* **4**, eaar7723 (2018).

37. 37.

Raichlen, D. A. & Gordon, A. D. Interpretation of footprints from Site S confirms human-like bipedal biomechanics in Laetoli hominins. *J. Hum. Evol.* **107**, 134–138 (2017).

38. 38.

Bennett, M. R. & Morse, S. A. in *Human Footprints: Fossilised Locomotion?* 101–136 (Springer, 2014).

39. 39.

Crompton, R. H. et al. Human-like external function of the foot, and fully upright gait, confirmed in the 3.66 million year old Laetoli hominin footprints by topographic statistics, experimental footprint-formation and computer simulation. *J. R. Soc. Interface* **9**, 707–719 (2012).

40. 40.

Meldrum, D. J., Lockley, M. G., Lucas, S. G. & Musiba, C. Ichnotaxonomy of the Laetoli trackways: the earliest hominin footprints. *J. Afr. Earth. Sci.* **60**, 1–12 (2011).

41. 41.

Ward, C. V., Kimbel, W. H. & Johanson, D. C. Complete fourth metatarsal and arches in the foot of *Australopithecus afarensis*. *Science* **331**, 750–753 (2011).

42. 42.

Hatala, K. G., Demes, B. & Richmond, B. G. Laetoli footprints reveal bipedal gait biomechanics different from those of modern humans and chimpanzees. *Proc. R. Soc. Lond. B* **283**, 20160235 (2016).

43. 43.

Leakey, M. G. et al. New hominin genus from eastern Africa shows diverse middle Pliocene lineages. *Nature* **410**, 433–440 (2001).

44. 44.

Haile-Selassie, Y., et al New species from Ethiopia further expands Middle Pliocene hominin diversity. *Nature* **521**, 483–488 (2015).

45. 45.

Musiba, C., et al Newly discovered hominin remains from upper Laetoli beds in northern Tanzania. Abstracts of the PaleoAnthropology Society, 2010 Meetings. *PaleoAnthropology* **2010**, A0001–A0040 (2010).

46. 46.

Pelissero, A. J. et al. Photogrammetry imaging: a fresh look at the Laetoli hominin footprints in relation to recent discoveries. *Am. J. Phys. Anthropol.* **162**, 312 (2017).

47. 47.

Pelissero, A., Musiba, C., Magori, C., Mwankunda, J. & Bunn, H. Rethinking the Laetoli hominin trackmakers in light of new prints at Site S. In *Paleoanthropology Society Meeting, Albuquerque, NM* (2019).

48. 48.

DeSilva, J., McNutt, E., Benoit, J. & Zipfel, B. One small step: a review of Plio-Pleistocene hominin foot evolution. *Am. J. Phys. Anthropol.* **168**, 63–140 (2019).

49. 49.

Musiba, C. M., Mabula, A., Selvaggio, M. & Magori, C. C. Pliocene animal trackways at Laetoli: research and conservation potential. *Ichnos* **15**, 166–178 (2008).

50. 50.

Bennett, M. R. et al. Early hominin foot morphology based on 1.5-million-year-old footprints from Ileret, Kenya. *Science* **323**, 1197–1201 (2009).

51. 51.

Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).

52. 52.

D'Aout, K., Meert, L., Van Gheluwe, B., De Clercq, D. & Aerts, P. Experimentally generated footprints in sand: analysis and consequences for the interpretation of fossil and forensic footprints. *Am. J. Phys. Anthropol.* **141**, 515–525 (2010).

53. 53.

Budka, M. et al. DigTrace Pro v.1.8.1 (Bournemouth University, 2016).

54. 54.

Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

55. 55.

Wickham, H., François, R., Henry, L. & Müller, K. dplyr: a grammar of data manipulation. Version 0.8.3 (2019).

56. 56.

Wickham, H. Reshape2: flexibly reshape data: a reboot of the Reshape package. Version 1.4.4 (CRAN, 2019).

57. 57.

Nagi, R. Maintaining detail and color definition when integrating color and grayscale rasters using no alteration of grayscale or intensity (NAGI) fusion method. In *Proc. 2012 AutoCarto* (2012).

## Acknowledgements

We thank R. Tuttle for his mentorship and inspiration; M. Larkin for her help collecting bear footprint data; R. Leakey, B. Wood, H. de Lumley, T. White, R. Crompton, R. Clarke, W. Harcourt-Smith, C. Stringer, H. Bonney and the London Natural History Museum, and J. Kibii and the National Museum of Kenya for assistance as we attempted to track down photographs and casts of the Laetoli A footprints; J. Sappington for photographic support; and V. Rossi and A. Metallo for help with 3D visualization; the Director General of Tanzania Commission for Science and Technology (COSTECH) for granting our team the COSTECH Research Permits (2019-370- NA-2019-2016); F. Manongi, J. Mwankunda and M. Mwambungu for allowing us to camp and work at Laetoli. Funding for this research was provided by the National Geographic (8748-10), Leakey Foundation (71483-001), National Science Foundation (BCS-1128170, DGE-080163, BCE-1730822, GRFP (No. 1840344), the Smithsonian Institution's Human Origins Program, the Evolving Earth Foundation, the Explorers Club, and the Claire Garber Goodman Fund-Dartmouth College.

## Author information

### Affiliations

1. Department of Biomedical Sciences, Ohio University Heritage College of Medicine, Athens, OH, USA

Ellison J. McNutt

2. Department of Integrative Anatomical Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Ellison J. McNutt & Elizabeth Kim

3. Department of Biology, Chatham University, Pittsburgh, PA, USA

Kevin G. Hatala & Erin Marie Williams-Hatala

4. Ecology, Evolution, Environment and Society Graduate Program, Dartmouth College, Hanover, NH, USA

Catherine Miller, Jesse Casana, Nathaniel J. Dominy, Luke D. Fannin & Jeremy M. DeSilva

5. Department of Anthropology, Dartmouth College, Hanover, NH, USA

Catherine Miller, Jesse Casana, Nathaniel J. Dominy, Luke D. Fannin, Austin C. Hill, Camille Johnson, Anjali Prabhat & Jeremy M. DeSilva

6. Dartmouth Library, Dartmout College, Hanover, NH, USA

James Adams

7. Information, Technology, and Consulting, Dartmouth College, Hanover, NH, USA

James Adams & Stephen Gaughan

8. Department of Anatomy, Cell Biology and Physiology, Indiana University School of Medicine, Indianapolis, IN, USA

Andrew S. Deane

9. Department of Cultural Heritage, Ngorongoro Conservation Area Authority, Arusha, Tanzania

Kallisti Fabian, Josephat Gurtu & Said Kallindo

10. Department of Occupational Therapy, Boston University, Boston, MA, USA

Simone V. Gill

11. Department of Anthropology, University of Colorado, Denver, CO, USA  
Ellie Gustafson, Rebeca Thornburg & Charles M. Musiba
12. Department of Anthropology, University of Pennsylvania, Philadelphia, PA, USA  
Austin C. Hill
13. Kilham Bear Center, Lyme, NH, USA  
Benjamin Kilham & Phoebe Kilham
14. Department of Geological and Environmental Sciences, Appalachian State University, Boone, NC, USA  
Cynthia Liutkus-Pierce & Brian Zimmer
15. Department of Anatomy, Idaho College of Osteopathic Medicine, Meridian, ID, USA  
Blaine Maley
16. Department of Anthropology, University College London, London, UK  
John Reader
17. Department of Anthropology, Napa Valley College, Napa, CA, USA  
Shirley Rubin
18. Department of Anatomy, NYIT College of Osteopathic Medicine, Old Westbury, NY, USA  
Nathan E. Thompson
19. Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa  
Charles M. Musiba & Jeremy M. DeSilva
20. Instituto Superior Politécnico de Tecnologia e Ciências, Luanda Angola, Angola  
Charles M. Musiba

## **Contributions**

E.J.M. contributed to study generation and design, data collection on bears and chimpanzees and analyses, and wrote the manuscript. K.G.H. contributed to data collection on humans and chimpanzees and analyses, and assisted with the manuscript. C.M. contributed to data collection during fieldwork at Laetoli and analyses. J.C. and A.C.H. contributed the photogrammetry analysis. J.A., A.C.H. and S.G. developed the contour maps of the footprints. A.S.D., K.F., L.D.F., J.G., E.G., S.K., B.M., A.P., S.R., R.T. and C.M.M. contributed to data collection during fieldwork at Laetoli. N.J.D. contributed to study design and assisted with the manuscript. S.V.G., J.M.D., K.G.H. and E.M.W.-H. contributed data collection on humans. J.M.D. also contributed to data collection during fieldwork at Laetoli, study design, and assisted with the manuscript. C.J. contributed to data collection on bears and analyses. B.K. and P.K. contributed to access to wild bears and data collection. C.L-P. and B.Z. contributed to access and data collection on the Engare Sero footprints. E.K. contributed to analyses of the Engare Sero footprints. J.R. contributed original photography of Laetoli Site A. N.E.T. contributed to data collection on chimpanzees. All authors contributed editorial comments to the manuscript.

## Corresponding author

Correspondence to [Ellison J. McNutt](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Laetoli print A3 and erosion.](#)

Photographs were taken by J. Reader (left: 1977) and J. DeSilva (right: 2019) at similar overhead angles and times of day (see similarities in shadows cast across the print). Notice the removed matrix infill in the hallucial impression and the presence of the previously unseen second digit impression in 2019 image. Below: comparisons of

the photogrammetric meshes of a pothole and footprints A2 and A3 using Cloud Compare<sup>15</sup>. The impact of erosion on the morphology of the footprints was assessed by quantifying changes to an oval depression located west of A3. Notice that significant changes occurred around the rim of the depression, as would be expected through erosion, that are absent around the rims of the A2 and A3. Instead, significant differences between A2 and A3 are located internally and are a result of a more thorough excavation of the prints.

### **Extended Data Fig. 2 Details of the best preserved Laetoli A hominin footprint (A3).**

Left image shows original 3D scan. Right image highlights the proposed impressions for the toes (blue circles) and matrix infill (green), as well as the potential evidence for midfoot mobility (red).

### **Extended Data Fig. 3 Incidence of bipedalism in *Ursus americanus* and Examples of kinematic data collection.**

(top) Pie charts showing the frequency of bipedal behaviors in wild *Ursus americanus*. The blue chart represents the time spent independently bipedal (locomotor or postural) out of the total 50.9 h in which bear behavior was observed. The orange chart represents the breakdown of time spent independently bipedal into its postural and locomotor (i.e., bears took one or more steps) components. The green chart represents the frequency of occurrences where bears walked 4 or more steps reflecting a similar circumstance to Laetoli trackway A. (bottom left) Juvenile female walks bipedally, unassisted through mud trackway. (bottom right) Example left footprint from one of the juvenile males.

[Source data](#)

### **Extended Data Fig. 4 Foot and gait comparisons across all comparative species.**

(a) Forefoot width to step length; (b) heel width to step length; (c) heel width to forefoot width; and (d) stride width to step length. Notice that for foot proportions (c) and stride width (d), Laetoli A is unlike the tracks produced by bears. In (d) the negative value in Laetoli A represents the fact that the track is demonstrating cross-stepping. (a-c) Boxplot represents median (center line), upper and lower quartiles (box limits), range (whiskers), and outliers (points) and individual footprints sample sizes for each species are indicated in the figure panels. (a-d) In all plots, n = 1 for all

Laetoli trackways, and chimpanzees were traveling bipedally. Data sources match Extended Data Table 1.

[Source data](#)

**Extended Data Fig. 5 Evidence challenging ursid hypothesis.**

(a) Ratio between toe impression dimensions across footprints from the comparative species and Laetoli A3. Values in blue=the ratio between the width of the 2<sup>nd</sup> digit compared to the hallux. Values in orange=ratio between the width of the 4<sup>th</sup> digit compared to the 5<sup>th</sup> in *Ursus americanus*. (Insert image) Comparison between 3-D scan of right ursid footprint (blue) and A3 (orange). Note the large size of the 5<sup>th</sup> digit impression in ursids but overall shape difference between the ursid track and A3. (b) Ratio of stride width to step length across the different species, including the Laetoli bipedal trackways. Boxplot definitions are as in Extended Data Fig. 4 and individual footprint sample sizes are indicated in the figure panels. In 5a, the chimpanzees were traveling quadrupedally; in 5b, chimpanzees were moving bipedally. Data sources match Extended Data Table 1.

[Source data](#)

**Extended Data Fig. 6 Boxplots showing the divergence ratio between the first digit and 2<sup>nd</sup> digit across the comparative species and two Laetoli trackways.**

Divergence ratio was measured by dividing the linear distance between the midpoint of digits one and two by the foot length. Chimpanzees were traveling quadrupedally and data were obtained from plantar pressure impressions. Boxplot definitions are as in Extended Data Fig. 4 and individual footprint sample sizes are indicated in the figure panel. Data sources match Extended Data Table 1. We attempted to measure hallucial divergence following Bennett et al<sup>50</sup>, but were stymied by matrix obscuring the deepest region of the ball of the foot.

[Source data](#)

**Extended Data Fig. 7 Histogram of Mahalanobis distances between the mean unshod human footprint and resampled unshod human footprints (gray) and chimpanzee footprints (yellow).**

As in Figure 3c, blue, orange, and green dashed lines represent samples from Laetoli G1, S1, and A, respectively. Only Laetoli A is labeled, for clarity. Sample sizes for

these samples as in Fig. 2c as well. The black dotted-dashed lines have been added to represent the average cross-stepping footprints produced by 10 adult habitually shod humans. All fall squarely within the distribution of unshod human footprints (probabilities of sampling tracks like these range from 0.20 to 0.94), and a great distance apart from the Laetoli A sample (green). Human cross-stepping footprints tended to be slightly closer to the human mean than the Laetoli S1 and G1 samples, but their distribution does overlap with the Laetoli S1 sample and with some of the Laetoli G1 samples. Cross-stepping footprints fell, on average, a Mahalanobis distance of 27.2 farther from the human mean than their “normal” walking counterpart.

### [Source data](#)

### [Extended Data Fig. 8 3D scans of Laetoli A footprints and their contours.](#)

Left panel: Complete scan of Laetoli A trackway with A1 at the bottom and A5 at the top. Right side: Zoomed in images of the individual footprints and their corresponding contour images.

### **Extended Data Table 1 Summary of average bipedal footprint metric data**

## **Supplementary information**

### [Supplementary Information](#)

This file contains Supplementary Discussion (Laetoli A rediscovery and description; *Agriotherium*) and Supplementary Methods (human cross-stepping footprint experiment; proportional toe depth measurement).

### [Reporting Summary](#)

### [Peer Review File](#)

## **Source data**

### [Source Data Fig. 2](#)

### [Source Data Extended Data Fig. 3](#)

### [Source Data Extended Data Fig. 4](#)

## Source Data Extended Data Fig. 5

## Source Data Extended Data Fig. 6

## Source Data Extended Data Fig. 7

## Source Data Extended Data Table 1

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## Reprints and Permissions

## About this article

### Cite this article

McNutt, E.J., Hatala, K.G., Miller, C. *et al.* Footprint evidence of early hominin locomotor diversity at Laetoli, Tanzania. *Nature* **600**, 468–471 (2021).  
<https://doi.org/10.1038/s41586-021-04187-7>

- Received: 24 March 2021
- Accepted: 22 October 2021
- Published: 01 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04187-7>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*Hominin footprints at Laetoli reveal a walk on the wild side\*\*](#)

- Stephanie M. Melillo

*Nature* (2021)

### [\*\*Hominin footprints at Laetoli reveal a walk on the wild side\*\*](#)

- Stephanie M. Melillo

News & Views 01 Dec 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04187-7>

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 08 July 2021](#)

# Mapping the human genetic architecture of COVID-19

- [COVID-19 Host Genetics Initiative](#)

*Nature* volume 600, pages 472–477 (2021)

- 148k Accesses
- 50 Citations
- 1717 Altmetric
- [Metrics details](#)

## Subjects

- [Genetics](#)
- [Genome-wide association studies](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

## Abstract

The genetic make-up of an individual contributes to the susceptibility and response to viral infection. Although environmental, clinical and social factors have a role in the chance of exposure to SARS-CoV-2 and the severity of COVID-19<sup>1,2</sup>, host genetics may also be important. Identifying host-specific genetic factors may reveal biological mechanisms of therapeutic relevance and clarify causal relationships of modifiable environmental risk factors for SARS-CoV-2 infection and outcomes. We formed a global network of researchers to investigate the role of human genetics in SARS-CoV-2 infection and COVID-19 severity. Here we describe the results of three genome-wide association meta-analyses that consist of up to 49,562 patients with COVID-19 from 46 studies across 19 countries. We report 13 genome-wide significant loci that

are associated with SARS-CoV-2 infection or severe manifestations of COVID-19. Several of these loci correspond to previously documented associations to lung or autoimmune and inflammatory diseases<sup>3,4,5,6,7</sup>. They also represent potentially actionable mechanisms in response to infection. Mendelian randomization analyses support a causal role for smoking and body-mass index for severe COVID-19 although not for type II diabetes. The identification of novel host genetic factors associated with COVID-19 was made possible by the community of human genetics researchers coming together to prioritize the sharing of data, results, resources and analytical frameworks. This working model of international collaboration underscores what is possible for future genetic discoveries in emerging pandemics, or indeed for any complex human disease.

[Download PDF](#)

## Main

The COVID-19 pandemic, caused by infection with SARS-CoV-2, has resulted in an enormous health and economic burden worldwide. One of the most remarkable features of SARS-CoV-2 infection is the variation in consequences, which range from asymptomatic to life-threatening, viral pneumonia and acute respiratory distress syndrome<sup>8</sup>. Although established host factors correlate with disease severity (for example, increasing age, being a man and higher body-mass index<sup>1</sup>), these risk factors alone do not explain all of the variability in disease severity observed across individuals.

Genetic factors contributing to COVID-19 susceptibility and severity may provide new biological insights into disease pathogenesis and identify mechanistic targets for therapeutic development or drug repurposing, as treating the disease remains a highly important goal despite the recent development of vaccines. Further supporting this line of inquiry, rare loss-of-function variants in genes involved in the type I interferon response may be involved in severe forms of COVID-19<sup>9,10,11</sup>. At the same time, several genome-wide association studies that investigate the contribution of common genetic variation<sup>12,13,14,15</sup> to COVID-19 have provided robust support for the involvement of several genomic loci associated with COVID-19 severity and susceptibility, with the strongest and most robust finding for severity being at the 3p21.31 locus<sup>12,13,14,15,16</sup>. However, much remains unknown about the genetic basis of susceptibility to SARS-CoV-2 and severity of COVID-19.

The COVID-19 Host Genetics Initiative (COVID-19 HGI) (<https://www.covid19hg.org/>)<sup>17</sup> is an international, open-science collaboration to share scientific methods and resources with research groups across the world with the goal to robustly map the host genetic determinants of SARS-CoV-2 infection and the

severity of the resulting COVID-19 disease. Here, we report the latest results of meta-analyses of 46 studies from 19 countries (Fig. 1) for COVID-19 host genetic effects.

**Fig. 1: Geographical overview of the contributing studies to the COVID-19 HGI and composition by major ancestry groups.**

 figure1

Populations are defined as African (AFR), admixed American (AMR), East Asian (EAS), European (EUR), Middle Eastern (MID) and South Asian (SAS).

## Meta-analyses of COVID-19

Overall, the COVID-19 HGI combined genetic data from 49,562 cases and 2 million controls across 46 distinct studies (Fig. 1). The data included studies from populations of different genetic ancestries, including European, admixed American, African, Middle Eastern, South Asian and East Asian individuals (Supplementary Table 1). An overview of the study design is provided in Extended Data Fig. 1. We performed case-control meta-analyses in three main categories of COVID-19 disease according to predefined and partially overlapping phenotypic criteria. These included (1) critically ill cases of COVID-19 defined as those individuals who required respiratory support in hospital or who died due to the disease; (2) cases of moderate or severe COVID-19 defined as those participants who were hospitalized due to symptoms associated with the infection; and (3) all cases with reported SARS-CoV-2 infection regardless of symptoms (Methods). Controls for all three analyses were selected as genetically ancestry-matched samples without known SARS-CoV-2 infection, if that information was available (Methods). The average age of the participants with COVID-19 across studies was 55 years (Supplementary Table 1). We report quantile-quantile plots in

Supplementary Fig. 1 and ancestry principal component plots for contributing studies in Extended Data Fig. 2.

Across our three analyses, we reported a total of 13 independent genome-wide significant loci associated with COVID-19 (the threshold of  $P < 1.67 \times 10^{-8}$  is adjusted for multiple trait testing) (Supplementary Table 2), most of which were shared between two or more COVID-19 phenotypes. Two of these loci are in very close proximity within the 3p21.31 region, which was previously reported as a single locus associated with COVID-19 severity<sup>12,13,14,15,16</sup> (Extended Data Fig. 3). Overall, we find six genome-wide significant associations for critical illness due to COVID-19, using data from 6,179 cases and 1,483,780 controls from 16 studies (Extended Data Fig. 4). Nine genome-wide significant loci were detected for moderate to severe hospitalized COVID-19 (including five of the six critical illness loci) from an analysis of 13,641 cases of COVID-19 and 2,070,709 controls across 29 studies (Fig. 2a, top). Finally, seven loci reached genome-wide significance in the analysis using data for all available 49,562 reported cases of SARS-CoV-2 infection and 1,770,206 controls, using data from a total of 44 studies (Fig. 2a, bottom). The proportion of cases with non-European genetic ancestry for each of the three analyses was 23%, 29% and 22%. We report the results for the lead variants at the 13 loci in different ancestry-group meta-analyses in Supplementary Table 3. We note that two loci, tagged by lead variants rs1886814 and rs72711165, had higher allele frequencies in southeast Asian (rs1886814; 15%) and East Asian genetic ancestry (rs72711165; 8%) whereas the minor allele frequencies in European populations were less than 3%. This highlights the value of including data from diverse populations for genetic discovery. We discuss the replication of previous findings and the new discoveries from these three analyses in the [Supplementary Note](#).

**Fig. 2: Genome-wide association results for COVID-19.**

---

 **figure2**

---

**a**, Top, results of a genome-wide association study of hospitalized cases of COVID-19 ( $n = 13,641$  cases and  $n = 2,070,709$  controls). Bottom, the results of reported SARS-CoV-2 infections ( $n = 49,562$  cases and  $n = 1,770,206$  controls). Loci highlighted in yellow (top) represent regions associated with the severity of the COVID-19 manifestation—that is, increased odds of more severe COVID-19 phenotypes. Loci highlighted in green (bottom) are regions associated with susceptibility to a SARS-CoV-2 infection—that is, the effect is the same across mild and severe COVID-19 phenotypes. We highlight in red genome-wide significant variants that had high heterogeneity across contributing studies and that were therefore excluded from the list of loci found. **b**, Results of gene prioritization using different evidence measures of gene annotation. Genes in the LD region, genes with coding variants and eGenes (fine-mapped *cis*-eQTL variant PIP > 0.1 in GTEx Lung) are annotated if in LD with a COVID-19 lead variant ( $r^2 > 0.6$ ). V2G, highest gene prioritized by the V2G score of Open Target Genetics.

## Variant effects on severity and susceptibility

We found no genome-wide significant sex-specific effects at the 13 loci. However, we did identify significant heterogeneous effects ( $P < 0.004$ ) across studies for 3 out of the 13 loci (Methods), which probably reflects the differential ascertainment of cases (Supplementary Table 2). There was a small number of overlapping samples ( $n = 8,380$  European ancestry;  $n = 745$  East Asian ancestry) between controls from the genOMICC and the UK Biobank studies, but leave-one-out sensitivity analyses did not reveal any bias in the corresponding effect sizes or  $P$  values (Extended Data Fig. 5 and [Supplementary Information](#)).

We next wanted to better understand whether the 13 significant loci were acting through mechanisms that increased the susceptibility to infection or that affected the progression of symptoms towards more severe disease. For all 13 loci, we compared the lead variant (strongest association *P* value) odds ratios (ORs) for the risk-increasing allele across our different COVID-19 phenotype definitions.

Focusing on the two better powered analyses: all cases with a reported SARS-CoV-2 infection and all cases hospitalized due to COVID-19, we find that four of the loci have similar odds ratios between these two analyses (Methods and Supplementary Table 2). Such consistency suggests a stronger link to susceptibility to SARS-CoV-2 infection rather than to the development of severe COVID-19. The strongest susceptibility signal was the previously reported *ABO* locus (rs912805253)<sup>12,13,15,16</sup>. Notably, and in agreement with a previously reported study<sup>15</sup>, we also report a locus within the 3p21.31 region that was more strongly associated with susceptibility to SARS-CoV-2 than progression to more severe COVID-19 phenotypes. rs2271616 showed a stronger association with a reported SARS-CoV-2 infection ( $P = 1.79 \times 10^{-34}$ ; OR (95% confidence interval (CI)) = 1.15 (1.13–1.18)) than hospitalization ( $P = 1.05 \times 10^{-5}$ ; OR (95% CI) = 1.12 (1.06–1.19)). For this locus—which contains additional independent signals—the linkage-disequilibrium (LD) pattern is discordant with the *P*-value expectation (Extended Data Fig. 6 and [Supplementary Note](#)), pointing to a key missing causal variant or to a potentially undiscovered multi-allelic or structural variant in this locus.

By contrast, 9 out of the 13 loci were associated with increased risk of severe symptoms with significantly larger odds ratios for hospitalized COVID-19 compared with the mildest phenotype of reported SARS-CoV-2 infection (eight loci were below the threshold of  $P < 0.004$  (test for effect size difference) and, in addition, the lead variant rs10774671 had a clear increase in odds ratios despite not passing this threshold) (Supplementary Table 2). We further compared the odds ratios for these nine loci for critical illness due to COVID-19 versus hospitalized due to COVID-19, and found that these loci exhibited a general increase in effect risk for critical illness (Methods, Extended Data Fig. 7a and Supplementary Table 4), but the lower power for association analysis of critically ill COVID-19 means that these results should be considered as suggestive. Overall, these results indicated that these nine loci were more likely to be associated with progression of the disease and worse outcome from SARS-CoV-2 infection compared to being associated with susceptibility to SARS-CoV-2 infection.

For some of these analyses, the controls were simply existing population controls without knowledge of SARS-CoV-2 infection or COVID-19 status, which may bias effect size estimates as some of these individuals may have either become infected with SARS-CoV-2 or developed COVID-19. We perform several sensitivity analyses

(Extended Data Fig. 7b, [Supplementary Note](#) and Supplementary Table 4) in which we show that using population controls can be a valid and powerful strategy for host genetic discovery of infectious disease, and particularly those that are widespread and with rare severe outcomes.

## Gene prioritization and association with other traits

To better understand the potential biological mechanism of each locus, we applied several approaches to prioritize candidate causal genes and explore additional associations with other diseases and traits. Of the 13 genome-wide significant loci, we found that nine loci implicated biologically plausible genes (Supplementary Tables 2, 5). Protein-altering variants in LD with lead variants implicated genes at six loci, including *TYK2* (chromosome and cytogenetic band (chr.) 19p13.2) and *PPPIR15A* (chr. 19q13.33). The COVID-19 lead variant rs74956615T>A in *TYK2*, which confers risk for critical illness (OR (95% CI) = 1.43 (1.29–1.59),  $P = 9.71 \times 10^{-12}$ ) and hospitalization due to COVID-19 (OR (95% CI) = 1.27 (1.18–1.36),  $P = 5.05 \times 10^{-10}$ ) is correlated with the missense variant rs34536443:G>C (p.Pro1104Ala;  $r^2 = 0.82$ ). This is consistent with the primary immunodeficiency described with complete *TYK2* loss of function<sup>3</sup> as this variant is known to reduce function<sup>18,19</sup>. By contrast, this missense variant was previously reported to be protective against autoimmune diseases (Extended Data Fig. 8 and Supplementary Table 6), including rheumatoid arthritis (OR = 0.74,  $P = 3.0 \times 10^{-8}$ ; UK Biobank SAIGE) and hypothyroidism (OR = 0.84,  $P = 1.8 \times 10^{-10}$ ; UK Biobank). At the 19q13.33 locus, the lead variant rs4801778, which was significantly associated with a reported SARS-CoV-2 infection (OR (95% CI) = 0.95 (0.93–0.96),  $P = 2.1 \times 10^{-8}$ ), is in LD ( $r^2 = 0.93$ ) with a missense variant rs11541192:G>A (p.Gly312Ser) in *PPPIR15A*.

A lung-specific *cis*-expression quantitative trait loci (*cis*-eQTLs) from GTEx v.8<sup>20</sup> ( $n = 515$ ) and the Lung eQTL Consortium<sup>21</sup> ( $n = 1,103$ ) provided further support for a subset of loci (Supplementary Table 7), including *FOXP4* (chr. 6p21.1) and *ABO* (chr. 9q34.2), *OASI/OAS3/OAS2* (chr. 12q24.13) and *IFNAR2/IL10RB* (21q22.11), where the COVID-19-associated variants modify gene expression in lung. Furthermore, our phenome-wide association study (PheWAS) analysis (Supplementary Table 6) implicated three additional loci related to lung function, with modest lung eQTL evidence—that is, the lead variant was not fine-mapped but significantly associated. An intronic variant rs2109069:G>A in *DPP9* (chr. 19p13.3), which is positively associated with critical illness, was previously reported to be risk-increasing for interstitial lung disease (tag lead variant rs12610495:A>G (p.Leu8Pro); OR = 1.29,  $P = 2.0 \times 10^{-12}$ )<sup>5</sup>. The COVID-19 lead variant rs1886814:A>C in the *FOXP4* locus is correlated ( $r^2 = 0.64$ ) with a lead variant of lung adenocarcinoma (tag variant is rs7741164; OR = 1.2,  $P = 6.0 \times 10^{-13}$ )<sup>6,22</sup> and similarly with a lead variant reported

for subclinical interstitial lung disease<sup>23</sup>. In severe COVID-19, lung cancer and interstitial lung disease, the minor, expression-increasing allele is associated with increased risk. We also found that intronic variants (chr. 1q22) and rs1819040:T>A in *KANSL1* (chr. 17q21.31), associated with protection against hospitalization due to COVID-19, were previously reported for reduced lung function (for example, tag lead variant rs141942982:G>T; OR (95% CI) = 0.96 (0.95–0.97),  $P = 1.00 \times 10^{-20}$ )<sup>7</sup>. Notably, the 17q21.31 locus is a well-known locus for structural variants containing a megabase inversion polymorphism (H1 and inverted H2 forms) and complex copy-number variations, in which the inverted H2 forms were shown to be positively selected in European individuals<sup>24,25</sup>.

Lastly, there are two loci in the 3p21.31 region with varying genes prioritized by different methods for different independent signals. For the severity lead variant rs10490770:T>C, we prioritized *CXCR6* with the Variant2Gene (V2G) algorithm<sup>26</sup>, although *LZTFL1* is the closest gene. The *CXCR6* has a role in chemokine signalling<sup>27</sup> and *LZTFL1* has been implicated in lung cancer<sup>28</sup>. rs2271616:G>T, which is associated with susceptibility, tags a complex region including several independent signals ([Supplementary Note](#)) that are all located within the gene body of *SLC6A20*, which encodes a protein that is known to functionally interact with the SARS-CoV-2 receptor ACE2<sup>29</sup>. However, none of the lead variants in the 3p21.31 region has been previously associated with other traits or diseases in our PheWAS analysis. Although these results provide supporting *in silico* evidence for candidate causal gene prioritization, further functional characterization is needed. Detailed locus descriptions and LocusZoom plots are provided in [Supplementary Fig. 2](#).

## Polygenic architecture of COVID-19

To further investigate the genetic architecture of COVID-19, we used results from meta-analyses including samples from European ancestries (sample sizes are described in the Methods and [Supplementary Table 1](#)) to estimate the heritability explained by common single-nucleotide polymorphisms—that is, the proportion of variation in the two phenotypes that was attributable to common genetic variants—and to determine whether heritability of COVID-19 phenotypes was enriched in genes that were specifically expressed in certain tissues<sup>30</sup> from the GTEx dataset<sup>31</sup>. We detected low, but significant, heritability across all three analyses (<1% on observed scale, all  $P$  values were  $P < 0.0001$ ) ([Supplementary Table 8](#)). The values are low compared to previously published studies<sup>14</sup>, but may be explained by differences in the reported estimate scale (observed versus liability), the specific method used, disease-prevalence estimates, phenotypic differences between patient cohorts or ascertainment of controls. Despite the low reported values, we found that heritability of a reported SARS-CoV-2 infection was significantly enriched in genes that were specifically expressed in the

lung ( $P = 5.0 \times 10^{-4}$ ) (Supplementary Table 9). These findings, together with the genome-wide significant loci identified in the meta-analyses, suggest that there is a significant polygenic architecture that can be better leveraged with future, larger, sample sizes.

## Genetic correlation and Mendelian randomization

Genetic correlations ( $r_g$ ) between the three COVID-19 phenotypes was high, although lower correlations were observed between hospitalized COVID-19 and reported SARS-CoV-2 infection (critical illness versus hospitalized:  $r_g$  (95% CI) = 1.37 (1.08–1.65),  $P = 2.9 \times 10^{-21}$ ; critical illness versus reported SARS-CoV-2 infection,  $r_g$  (95% CI) = 0.96 (0.71–1.20),  $P = 1.1 \times 10^{-14}$ ; hospitalized versus reported SARS-CoV-2 infection:  $r_g$  (95% CI) = 0.85 (0.68–1.02),  $P = 1.1 \times 10^{-22}$ ). To better understand which traits are genetically correlated and/or potentially causally associated with COVID-19 hospitalization, critical illness and reported SARS-CoV-2 infection, we chose a set of 38 disease, health and neuropsychiatric phenotypes as potential COVID-19 risk factors based on their clinical correlation with disease susceptibility, severity or mortality (Supplementary Table 10).

We found evidence (false-discovery rate (FDR) < 0.05) of significant genetic correlations between nine traits and hospitalized COVID-19 and reported SARS-CoV-2 infection (Fig. 3, Extended Data Fig. 9 and Supplementary Table 11). Notably, genetic liability to ischaemic stroke was only significantly positively correlated with critical illness or hospitalization due to COVID-19, but not with a higher likelihood of reported SARS-CoV-2 infection (infection  $r_g$  = 0.019 versus hospitalization  $r_g$  = 0.41,  $z = 2.7$ ,  $P = 0.006$ ; infection  $r_g$  = 0.019 versus critical illness  $r_g$  = 0.40,  $z = 2.49$ ,  $P = 0.013$ ).

**Fig. 3: Genetic correlations and Mendelian randomization causal estimates between 38 traits and COVID-19 critical illness, hospitalization and reported SARS-CoV-2 infection.**

 figure3

Larger squares correspond to  $P$  values with higher significance, with genetic correlations ( $r_g$ ) or Mendelian randomization (MR) causal estimates significantly different from zero. The size of each coloured square indicates the magnitude of the  $P$  value, with  $P < 0.05$  shown as a full-sized square,  $P = 0.05–0.1$  as a large square,  $P = 0.1–0.5$  as a medium square and  $P > 0.5$  as a small square. Genetic correlations or causal estimates that are significantly different from zero at an FDR of 5% are marked with an asterisk. Two-sided  $P$  values were calculated using LDSC for genetic correlations and inverse-variance-weighted analysis for Mendelian randomization. ADHD, attention-deficit hyperactivity disorder; BMI, body mass index; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate.

We next used two-sample Mendelian randomization to infer potentially causal relationships between these traits. After correcting for multiple testing (FDR < 0.05), eight exposure–COVID-19 trait pairs showed suggestive evidence of a causal association (Fig. 3, Extended Data Fig. 10, Supplementary Table 12 and Supplementary Fig. 3). Five of these associations were robust to potential violations of the underlying assumptions of Mendelian randomization. Corroborating our genetic correlation results and evidence from epidemiological studies, genetically predicted higher body-mass index (OR (95% CI) = 1.4 (1.3–1.6),  $P = 8.5 \times 10^{-11}$ ) and smoking (OR (95% CI) = 1.9 (1.3–2.8),  $P = 0.0012$ ) were associated with increased risk of COVID-19 hospitalization, with body-mass index also being associated with increased risk of SARS-CoV-2 infection (OR (95% CI) = 1.1 (1.1–1.2),  $P = 4.8 \times 10^{-7}$ ). Genetically predicted increased height (OR (95% CI) = 1.1 (1–1.1)),  $P = 8.9 \times 10^{-4}$ ) was associated with an increased risk of reported SARS-CoV-2 infection, whereas a genetically predicted higher red-blood-cell count (OR (95% CI) = 0.93 (0.89–0.96),  $P = 5.7 \times 10^{-5}$ ) was associated with a reduced risk of reported SARS-CoV-2 infection. Despite evidence of a genetic correlation between type II diabetes and COVID-19 outcomes, there was no evidence of a causal association in the Mendelian randomization analyses, which suggests that the observed genetic correlations are due to pleiotropic effects between body-mass index and type 2 diabetes. Further sensitivity analyses relating to sample overlap are discussed in the [Supplementary Information](#).

## Discussion

The COVID-19 HGI has brought together investigators from across the world to advance genetic discovery for SARS-CoV-2 infection and severe COVID-19 disease. We report 13 genome-wide significant loci associated with some aspect of SARS-CoV-2 infection or COVID-19. Many of these loci overlap with previously reported associations with lung-related phenotypes or autoimmune or inflammatory diseases, but some loci have no obvious candidate gene.

Four out of the thirteen genome-wide significant loci showed similar effects in the reported SARS-CoV-2 infection analysis (a proxy for disease susceptibility) and all-hospitalized COVID-19 (a proxy for disease severity). Of these, one locus was in close proximity to, yet independent of, the major genetic signal for COVID-19 severity at the 3p21.31 locus. Notably, this locus was associated with COVID-19 susceptibility rather than severity. The locus overlaps *SLC6A20*, which encodes an amino acid transporter that interacts with ACE2. Nonetheless, we caution that more data are needed to resolve the nature of the relationship between genetic variation and COVID-19 at this locus, particularly as the physical proximity, LD structure and patterns of association suggest that untagged genetic variation could drive the association signal in the region. Our findings support the notion that some genetic variants, most notably at the *ABO* and *PPP1R15A* loci, in addition to *SLC6A20*, can indeed affect susceptibility to infection rather than progression to severe COVID-19 once infected.

Several of the loci reported here—as noted in previous publications<sup>12,14</sup>—intersect with well-known genetic variants that have established genetic associations. Examples of these include variants at *DPP9* and *FOXP4*, which show previous evidence of increasing risk for interstitial lung disease<sup>5</sup>, and missense variants within *TYK2* that show a protective effect on several autoimmune-related diseases<sup>32,33,34,35</sup>. Together with the heritability enrichment observed in genes expressed in lung tissues, these results highlight the involvement of lung-related biological pathways in the development of severe COVID-19. Several other loci show no previously documented genome-wide significant associations, despite the high significance and attractive candidate genes for COVID-19 (for example, *CXCR6*, *LZTFL1*, *IFNAR2* and *OAS1/OAS2/OAS3* loci). The previously reported associations for the strongest association for COVID-19 severity at the 3p21.31 locus and monocytes count are likely to be due to proximity and not a true co-localization.

Increasing the global representation in genetic studies enhances the ability to detect novel associations. Two of the loci that affect disease severity were only discovered by including the four studies of individuals with East Asian ancestry. One of these loci—close to *FOXP4*—is common particularly in East Asian participants (32%) as well as admixed American participants in the Americas (20%) and Middle Eastern participants (7%), but has a low frequency in most European ancestries (2–3%) in our data. Although we cannot be certain of the mechanism of action, the *FOXP4* association is an attractive biological target, as it is expressed in the proximal and distal airway epithelium<sup>36</sup> and has been shown to have a role in controlling epithelial cell fate during lung development<sup>37</sup>. The COVID-19 HGI continues to pursue expansion of the datasets included in the analyses of the consortium to populations from underrepresented populations in upcoming data releases. We plan to release ancestry-specific results in full once the sample sizes allow for a well-powered meta-analysis.

Care should be taken when interpreting the results from a meta-analysis because of challenges with case and control ascertainment and collider bias (see [Supplementary Note](#) for a more detailed discussion on study limitations). Drawing a comprehensive and reproducible map of the host genetics factors associated with COVID-19 severity and SARS-CoV-2 requires a sustained international effort to include diverse ancestries and study designs. To accelerate downstream research and therapeutic discovery, the COVID-19 HGI regularly publishes meta-analysis results from periodic data freezes on the website <https://www.covid19hg.org/> and provides an interactive explorer through which researchers can browse the results and the genomic loci in more detail. Future work will be required to better understand the biological and clinical value of these findings. Continued efforts to collect more samples and detailed phenotypic data should be endorsed globally, allowing for more thorough investigation of variable, heritable symptoms, particularly in light of the newly emerging strains of SARS-CoV-2, which may provoke different host responses that lead to disease.

## Methods

### Contributing studies

All of the participants were recruited following protocols approved by local Institutional Review Boards; this information is collected in Supplementary Table 1 for all 46 studies. All protocols followed local ethics recommendations and informed consent was obtained when required. Information about sample numbers, sex and age from for each contributing study is given in Supplementary Table 1. In total, 16 studies contributed data to the analysis of critical illness due to COVID-19, 29 studies contributed data to hospitalized COVID-19 analysis and 44 studies contributed to the analysis of all cases of COVID-19. Each individual study that contributed data to a particular analysis met a minimum threshold of 50 cases, as defined by the phenotypic criteria, for statistical robustness. The effective sample sizes for each ancestry group shown in Fig. 1 were calculated for display using the formula:  $((4 \times N_{\text{case}} \times N_{\text{control}})/(N_{\text{case}} + N_{\text{control}}))$ . Details of contributing research groups are provided in Supplementary Table 1.

### Phenotype definitions

COVID-19 disease status (critical illness and hospitalization status) was assessed following the Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia<sup>38</sup>. The critically ill COVID-19 group included patients who were hospitalized owing to symptoms associated with laboratory-confirmed SARS-CoV-2 infection and who required respiratory support or whose cause of death was associated with COVID-19.

The hospitalized COVID-19 group included patients who were hospitalized owing to symptoms associated with laboratory-confirmed SARS-CoV-2 infection.

The reported SARS-CoV-2 infection group included individuals with laboratory-confirmed SARS-CoV-2 infection or electronic health record, ICD coding or clinically confirmed COVID-19, or self-reported COVID-19 (for example, by questionnaire), with or without symptoms of any severity. Genetic-ancestry-matched control individuals for the three case definitions were sourced from population-based cohorts, including individuals whose exposure status to SARS-CoV-2 was either unknown or infection-negative for questionnaire/electronic-health-record-based cohorts. Additional information regarding individual studies contributing to the consortium are described in Supplementary Table 1.

## Genome-wide association studies and meta-analyses

Each contributing study genotyped the samples and performed quality controls, data imputation and analysis independently, but following the consortium recommendations (information is available at <https://www.covid19hg.org/>). We recommended that genome-wide association study (GWAS) analyses were run using Scalable and Accurate Implementation of GEneralized mixed model (SAIGE)<sup>39</sup> on chromosomes 1–22 and X. The recommended analysis tool was SAIGE, but studies also used other software such as PLINK<sup>40</sup>. The suggested covariates were age, age<sup>2</sup>, sex, age × sex and the 20 first principal components. Any other study-specific covariates to account for known technical artefacts could be added. SAIGE automatically accounts for sample relatedness and case–control imbalances. Quality-control and analysis approaches for individual studies are reported in Supplementary Table 1.

Study-specific summary statistics were then processed for meta-analysis. Potential false positives, inflation and deflation were examined for each submitted GWAS. Allele frequency plots against gnomAD 3.0 genomes were manually inspected for each study. Standard error values as a function of the effective sample size were used to find studies that deviated from the expected trend. Summary statistics passing this manual quality control were included in the meta-analysis. Variants with an allele frequency of >0.1% and an imputation INFO score of >0.6 were carried forward from each study. Variants and alleles were lifted over to genome build GRCh38, if needed, and harmonized to gnomAD 3.0 genomes<sup>41</sup> by finding matching variants by strand flipping or switching the ordering of alleles. If multiple matching variants were included, the best match was chosen according to the minimum fold change in absolute allele frequency. Meta-analysis was performed using the inverse-variance-weighted (IVW) method on variants that were present in at least two-thirds of the studies contributing to the phenotype analysis. The method summarizes effect sizes

across the multiple studies by computing the mean of the effect sizes weighted by the inverse variance in each individual study.

We report 13 meta-analysis variants that pass the genome-wide significance threshold after adjusting the threshold for multiple traits tested ( $P < 5 \times 10^{-8}/3$ ). We report the unadjusted  $P$  values for each variant. We tested for heterogeneity between estimates from contributing studies using Cochran's  $Q$ -test<sup>42,43</sup>. This is calculated for each variant as the weighted sum of squared differences between the effects sizes and their meta-analysis effect, the weights being the inverse variance of the effect size.  $Q$  is distributed as a  $\chi^2$  statistic with  $k$  (number of studies) minus one degrees of freedom. Two loci reached genome-wide significance but were excluded from the significant results in Supplementary Table 2 due to heterogeneity between estimates from contributing studies and missingness between studies at chr. 6: 31057940–31380334 and chr. 7: 54671568–54759789; however, these regions are not excluded from the corresponding summary statistics in data release 5 (COVID-19 HGI (<https://www.covid19hg.org/results/r5/>) and GWAS Catalog (study code GCST011074)). For each of the lead variants reported in Supplementary Table 2, we aimed to find loci specific to susceptibility or severity by testing whether there was heterogeneity between the effect sizes associated with hospitalized COVID-19 (progression to severe disease) and reported SARS-CoV-2 infection. We used the Cochran's  $Q$  measure<sup>42,43</sup>, calculated for each variant as the weighted sum of squared differences between the two analysis effect sizes and their meta-analysis effect with the weights being the inverse variance of the effect size. A significant  $P$  value of  $P < 0.004$  ((0.05/13 loci) for multiple tests) indicates that the effect sizes for a particular variant are significantly different in the two analyses (Supplementary Table 2). For the nine loci, in which the lead variant effect size was significantly higher for hospitalized COVID-19, we carried out the same test again but comparing effect sizes from hospitalized COVID-19 with critically ill COVID-19 (Supplementary Table 4). Furthermore, we carried out the same test comparing meta-analysed hospitalized COVID-19 (population as controls) and hospitalized COVID-19 (SARS-CoV-2-positive but non-hospitalized as controls) (Supplementary Table 4). For these pairs of phenotype comparisons, we generated new meta-analysis summary statistics to use; including only those studies that could contribute data to both phenotypes that were under comparison.

## Principal component projection

To project every GWAS participant into the same principal component (PC) space, we used pre-computed PC loadings and reference allele frequencies. For reference, we used unrelated samples from the 1000 Genomes Project and the Human Genome Diversity Project and computed PC loadings and allele frequencies for the 117,221 single-nucleotide polymorphisms (SNPs) that (1) are available in every cohort; (2)

have a minor allele frequency of >0.1% in the reference; and (3) are LD-pruned ( $r^2 < 0.8$ ; 500-kb window). We then asked each cohort to project their samples using our automated script provided at <https://github.com/covid19-hg/>. It internally uses the PLINK2<sup>44</sup> --score function with the variance-standardize option and reference allele frequencies (--read-freq); so that each cohort-specific genotype/dosage matrix is mean-centred and variance-standardized with respect to reference allele frequencies, but not cohort-specific allele frequencies. We further normalized the projected PC scores by dividing the values by a square root of the number of variants used for projection to account for a subtle difference due to missing variants.

## Gene prioritization

To prioritize candidate causal genes reported in full in Supplementary Table 2, we used various gene prioritization approaches using both locus-based and similarity-based methods. Because we only describe the *in silico* gene prioritization results without characterizing the actual functional activity *in vitro* or *in vivo*, we aimed to provide a systematic approach to nominate potential causal genes in a locus using the following criteria.

- (1) The closest gene: a gene that is closest to a lead variant by distance to the gene body.
- (2) Genes in the LD region: genes that overlap with a genomic range containing any variants in LD ( $r^2 > 0.6$ ) with a lead variant. For LD computation, we retrieved LD matrices provided by gnomAD v.2.1.1<sup>41</sup> for each population analysed in this study (except for admixed American, Middle Eastern and South Asian genetic ancestry populations, for whom data are not available). We then constructed a weighted-average LD matrix by per-population sample sizes in each meta-analysis, which we used as a LD reference.
- (3) Genes with coding variants: genes with at least one loss-of-function or missense variant (annotated by VEP<sup>45</sup> v.95 with GENCODE v.29) that is in LD with a lead variant ( $r^2 > 0.6$ ).
- (4) eGenes: genes with at least one fine-mapped *cis*-eQTL variant (PIP > 0.1) that is in LD with a lead variant ( $r^2 > 0.6$ ) (Supplementary Table 5). We retrieved fine-mapped variants from the GTEx v.8<sup>20</sup> (<https://www.finucanelab.org/>) and eQTL catalogue<sup>46</sup>. In addition, we looked up significant associations in the Lung eQTL Consortium<sup>21</sup> ( $n = 1,103$ ) to further support our findings in lung with a larger sample size (Supplementary Table 7). We note that, in contrast to the GTEx or eQTL catalogue, we only looked at associations and did not fine-map our data to the Lung eQTL Consortium data.

(5) V2G: a gene with the highest overall V2G score based on Open Targets Genetics (OTG)<sup>26</sup>. For each variant, the overall V2G score aggregates differentially weighted evidence of variant–gene associations from several data sources, including molecular *cis*-QTL data (for example, *cis*-protein QTLs from ref. <sup>47</sup>, *cis*-eQTLs from GTEx v.7 and so on), interaction-based datasets (for example, promoter capture Hi-C), genomic distance and variant effect predictions (VEP) from Ensembl. A detailed description of the evidence sources and weights used is provided in the OTG documentation (<https://genetics-docs.opentargets.org/our-approach/data-pipeline>)<sup>26</sup>.

## Phenome-wide association study

To investigate the evidence of shared effects of 15 index variants for COVID-19 and previously reported phenotypes, we performed a phenome-wide association study. We considered phenotypes in OTG obtained from the GWAS catalogue (this included studies with and without full summary statistics,  $n = 300$  and 14,013, respectively)<sup>48</sup> and from the UK Biobank. Summary statistics for UK Biobank traits were extracted from SAIGE<sup>39</sup> for binary outcomes ( $n = 1,283$  traits) and Neale v.2 ( $n = 2,139$  traits) for both binary and quantitative traits (<http://www.nealelab.is/uk-biobank/>) and FinnGen Freeze 4 cohort ([https://www.finngen.fi/en/access\\_results](https://www.finngen.fi/en/access_results)). We report PheWAS results for phenotypes for which the lead variants were in high LD ( $r^2 > 0.8$ ) with the 13 genome-wide significant lead variants from our main COVID-19 meta-analysis (Supplementary Table 6). This conservative approach allowed spurious signals primarily driven by proximity rather than actual colocalization to be removed (see [Methods](#)).

To remove plausible spurious associations, we retrieved phenotypes for GWAS lead variants that were in LD ( $r^2 > 0.8$ ) with COVID-19 index variants.

## Heritability

LD score regression v.1.0.1<sup>49</sup> was used to estimate the SNP heritability of the phenotypes from the meta-analysis summary statistic files. As this method depends on matching the LD structure of the analysis sample to a reference panel, the summary statistics of European ancestry only were used. Sample sizes were  $n = 5,101$  critically ill cases of COVID-19 and  $n = 1,383,241$  control participants,  $n = 9,986$  hospitalized cases of COVID-19 and  $n = 1,877,672$  control participants, and  $n = 38,984$  cases and  $n = 1,644,784$  control participants for the analysis of all cases—all including the 23andMe cohort. Pre-calculated LD scores from the 1000 Genomes European reference population were obtained online (<https://data.broadinstitute.org/alkesgroup/LDScore/>). Analyses were conducted using the standard program settings for variant filtering (removal of non-HapMap3

SNPs, the HLA region on chromosome 6, non-autosomal,  $\chi^2 > 30$ , minor allele frequency of <1%, or allele mismatch with reference). We additionally report SNP heritability estimates for the all-ancestries meta-analyses, calculated using European panel LD scores, in Supplementary Table 8.

## Partitioned heritability

We used partitioned LD score regression<sup>50</sup> to partition COVID-19 SNP heritability in cell types in our summary statistics for European ancestry only. We ran the analysis using the baseline model LD scores calculated for European populations and regression weights that are available online (<https://github.com/bulik/ldsc>). We used the COVID-19 summary statistics for European ancestry only for the analysis.

## Genome-wide association summary statistics

We obtained genome-wide association summary statistics for 43 complex-disease, neuropsychiatric, behavioural or biomarker phenotypes (Supplementary Table 10). These phenotypes were selected based on their putative relevance to COVID-19 susceptibility, severity or mortality, with 19 selected based on the Centers for Disease Control list of underlying medical conditions associated with COVID-19 severity<sup>51</sup> or traits reported to be associated with increased risk of COVID-19 mortality by OpenSafely<sup>52</sup>. Summary statistics generated from GWAS using individuals of European ancestry were preferentially selected if available. These summary statistics were used in subsequent genetic correlation and Mendelian randomization analyses.

## Genetic correlation

LD score regression<sup>50</sup> was also used to estimate the genetic correlations between our COVID-19 meta-analysis phenotypes reported using samples of only European ancestry, and between these and the curated set of 38 summary statistics. Genetic correlations were estimated using the same LD score regression settings as for heritability calculations. Differences between the observed genetic correlations of SARS-CoV-2 infection and COVID-19 severity were compared using a z-score method<sup>53</sup>.

## Mendelian randomization

Two-sample Mendelian randomization was used to evaluate the potential for causal association of the 38 traits on COVID-19 hospitalization, on COVID-19 severity and reported SARS-CoV-2 infection using samples of only European ancestry. Independent genome-wide significant SNPs robustly associated with the exposures of

interest ( $P < 5 \times 10^{-8}$ ) were selected as genetic instruments by performing LD clumping using PLINK<sup>40</sup>. We used a strict  $r^2$  threshold of 0.001, a 10-Mb clumping window, and the European reference panel from the 1000 Genomes Project<sup>54</sup> to discard SNPs in LD with another variant with a smaller  $P$ -value association. For genetic variants that were not present in the hospitalized COVID-19 analysis, PLINK was used to identify proxy variants that were in LD ( $r^2 > 0.8$ ). Next, the exposure and outcome datasets were harmonized using the R package TwoSampleMR<sup>55</sup>. Namely, we ensured that the effect of a variant on the exposure and outcome corresponded to the same allele, we inferred positive-strand alleles and dropped palindromes with ambiguous allele frequencies, as well as incompatible alleles. Supplementary Table 10 includes the harmonized datasets used in the analyses.

The global test from Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO)<sup>56</sup> software was used to investigate overall horizontal pleiotropy. In brief, the standard IVW meta-analytic framework was used to calculate the average causal effect by excluding each genetic variant used to instrument the analysis. A global statistic was calculated by summing the observed residual sum of squares, that is, the difference between the effect predicted by the IVW slope excluding the SNP, and the observed effect of the SNP on the outcome. Overall horizontal pleiotropy was subsequently analysed by comparing the observed residual sum of squares, with the residual sum of squares expected under the null hypothesis of no pleiotropy. The MR-PRESSO global test was shown to perform well when the outcome and exposure GWASs are not disjoint (although the power to detect horizontal pleiotropy is slightly reduced by complete sample overlap). We also used the regression intercept in MR-Egger<sup>57</sup> to evaluate potential bias due to directional pleiotropic effects. This additional check was used in Mendelian randomization analyses with an  $\{I\}_{\{\backslash rm{GX}\}}^2$  index surpassing the recommended threshold ( $\{I\}_{\{\backslash rm{GX}\}}^2 > 90\%$ )<sup>58</sup>. Contingent on the MR-PRESSO global test results we analysed the causal effect of each exposure on COVID-19 hospitalization by using a fixed-effect IVW meta-analysis as the primary analysis, or, if pleiotropy was present, the MR-PRESSO outlier-corrected test. The IVW approach estimates the causal effect by aggregating the single-SNP causal effects (obtained using the ratio of coefficients method—that is, the ratio of the effect of the SNP on the outcome over the effect of the SNP on the exposure) in a fixed-effects meta-analysis. The SNPs were assigned weights based on their inverse variance. The IVW method confers the greatest statistical power for estimating causal associations<sup>59</sup>, but assumes that all variants are valid instruments and can produce biased estimates if the average pleiotropic effect differs from zero. Alternatively, when horizontal pleiotropy was present, we used the MR-PRESSO outlier-corrected method to correct the IVW test by removing outlier SNPs. We conducted further sensitivity analyses using alternative Mendelian randomization methods that provide consistent estimates of the causal effect even when some instrumental variables are invalid, at the cost of reduced

statistical power including: (1) Weighted median estimator (WME); (2) weighted mode-based estimator (WMBE); and (3) MR-Egger regression. Robust causal estimates were defined as those that were significant at an FDR of 5% and either (1) showed no evidence of heterogeneity (MR-PRESSO global test  $P > 0.05$ ) or horizontal pleiotropy (Egger intercept  $P > 0.05$ ); or (2) in the presence of heterogeneity or horizontal pleiotropy, the WME-, WMBE-, MR-Egger- or MR-PRESSO-corrected estimates were significant ( $P < 0.05$ ). All statistical analyses were conducted using R v.4.0.3. Mendelian randomization analysis was performed using the ‘TwoSampleMR’ v.0.5.5 package<sup>55</sup>.

## Website and data distribution

In anticipation of the need to coordinate many international partners around a single meta-analysis effort, we created the COVID-19 HGI website (<https://covid19hg.org>). We were able to centralize information, recruit partner studies, rapidly distribute summary statistics and present preliminary interpretations of the results to the public. Open meetings are held on a monthly basis to discuss future plans and new results; video recordings and supporting documents are shared (<https://covid19hg.org/meeting-archive>). This centralized resource provides a conceptual and technological framework for organizing global academic and industry groups around a shared goal. The website source code and additional technical details are available at <https://github.com/covid19-hg/covid19hg>.

To recruit new international partner studies, we developed a workflow in which new studies are registered and verified by a curation team (<https://covid19hg.org/register>). Users can explore the registered studies using a customized interface to find and contact studies with similar goals or approaches (<https://covid19hg.org/partners>). This helps to promote organic assembly around focused projects that are adjacent to the centralized effort (<https://covid19hg.org/projects>). Visitors can query study information, including study design and research questions. Registered studies are visualized on a world map and are searchable by institutional affiliation, city and country.

To encourage data sharing and other forms of participation, we created a rolling acknowledgements page (<https://covid19hg.org/acknowledgements>) and directions on how to contribute data to the central meta-analysis effort (<https://covid19hg.org/data-sharing>). Upon the completion of each data freeze, we post summary statistics, plots and sample size breakdowns for each phenotype and contributing cohort (<https://covid19hg.org/results>). The results can be explored using an interactive web browser (<https://app.covid19hg.org>). Several computational research groups carry out follow-up analyses, which are made available for download (<https://covid19hg.org/in-silico>). To enhance scientific communication to

the public, preliminary results are described in blog posts by the scientific communications team and shared on Twitter. The first post was translated to 30 languages with the help of 85 volunteer translators. We compile publications and preprints submitted by participating groups and summarize genome-wide significant findings from these publications (<https://covid19hg.org/publications>).

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Summary statistics generated by the COVID-19 HGI are available at <https://www.covid19hg.org/results/r5/> and are available in the GWAS Catalog (study code GCST011074). The analyses described here include the freeze-5 data. COVID-19 HGI continues to regularly release new data freezes. Summary statistics for non-European ancestry samples are not currently available due to the small individual sample sizes of these groups, but results for lead variants of 13 loci are reported in Supplementary Table 3. Individual level data can be requested directly from contributing studies, listed in Supplementary Table 1. We used publicly available data from GTEx (<https://gtexportal.org/home/>), the Neale lab (<http://www.nealelab.is/uk-biobank/>), Finucane lab (<https://www.finucanelab.org>), the FinnGen Freeze 4 cohort ([https://www.finngen.fi/en/access\\_results](https://www.finngen.fi/en/access_results)) and the eQTL catalogue release 3 (<http://www.ebi.ac.uk/eqtl/>).

## Code availability

The code for summary statistics lift-over, the projection PCA pipeline including precomputed loadings and meta-analyses are available on GitHub (<https://github.com/covid19-hg/>) and the code for the Mendelian randomization and genetic correlation pipeline is available on GitHub at <https://github.com/marcoralab/MRcovid>.

## References

1. 1.  
Docherty, A. B. et al. Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *Br. Med. J.* **369**, m1985 (2020).
2. 2.  
Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
3. 3.  
Dendrou, C. A. et al. Resolving *TYK2* locus genotype-to-phenotype differences in autoimmunity. *Sci. Transl. Med.* **8**, 363ra149 (2016).
4. 4.  
Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
5. 5.  
Fingerlin, T. E. et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* **45**, 613–620 (2013).
6. 6.  
Wang, Z. et al. Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women. *Hum. Mol. Genet.* **25**, 620–629 (2016).

7. 7.

Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).

8. 8.

Buitrago-Garcia, D. et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346 (2020).

9. 9.

van der Made, C. I. et al. Presence of genetic variants among young men with severe COVID-19. *J. Am. Med. Assoc.* **324**, 663–673 (2020).

10. 10.

Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).

11. 11.

Povysil, G. et al. Rare loss-of-function variants in type I IFN immunity genes are not associated with severe COVID-19. *J. Clin. Invest.* 147834 (2021).

12. 12.

Severe COVID-19 GWAS Group. Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).

13. 13.

Shelton, J. F. et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).

14. 14.

Pairo-Castineira, E. et al. Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).

15. 15.

Roberts, G. H. L. et al. AncestryDNA COVID-19 host genetic study identifies three novel loci. Preprint at <https://doi.org/10.1101/2020.10.06.20205864> (2020).

16. 16.

Kosmicki, J. A. et al. A catalog of associations between rare coding variants and COVID-19 outcomes. Preprint at <https://doi.org/10.1101/2020.10.28.20221804> (2021).

17. 17.

COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).

18. 18.

Couturier, N. et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* **134**, 693–703 (2011).

19. 19.

Li, Z. et al. Two rare disease-associated Tyk2 variants are catalytically impaired but signaling competent. *J. Immunol.* **190**, 2335–2344 (2013).

20. 20.

GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

21. 21.

Hao, K. et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8**, e1003029 (2012).

22. 22.

Dai, J. et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir. Med.* **7**, 881–891 (2019).

23. 23.

Manichaikul, A. et al. Genome-wide association study of subclinical interstitial lung disease in MESA. *Respir. Res.* **18**, 97 (2017).

24. 24.

Stefansson, H. et al. A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).

25. 25.

Boettger, L. M., Handsaker, R. E., Zody, M. C. & McCarroll, S. A. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat. Genet.* **44**, 881–885 (2012).

26. 26.

Ghoussaini, M. et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).

27. 27.

Xiao, G. et al. CXCL16/CXCR6 chemokine signaling mediates breast cancer progression by pERK1/2-dependent mechanisms. *Oncotarget* **6**, 14165–14178 (2015).

28. 28.

Wei, Q. et al. LZTFL1 suppresses lung tumorigenesis by maintaining differentiation of lung epithelial cells. *Oncogene* **35**, 2655–2663 (2016).

29. 29.

Vuille-dit-Bille, R. N. et al. Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids* **47**, 693–705 (2015).

30. 30.

Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

31. 31.

GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

32. 32.

Eyre, S. et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336–1340 (2012).

33. 33.

Tsoi, L. C. et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* **8**, 15382 (2017).

34. 34.

Langefeld, C. D. et al. Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 (2017).

35. 35.

Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

36. 36.

Lu, M. M., Li, S., Yang, H. & Morrisey, E. E. *Foxp4*: a novel member of the Foxp subfamily of winged-helix genes co-expressed with *Foxp1* and *Foxp2* in pulmonary and gut tissues. *Gene Expr. Patterns* **2**, 223–228 (2002).

37. 37.

Li, S. et al. Foxp1/4 control epithelial cell fate during lung development and regeneration through regulation of anterior gradient 2. *Development* **139**, 2500–2509 (2012).

38. 38.

Wei, P.-F. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7). *Chin. Med. J. (Engl.)* **133**, 1087–1095 (2020).

39. 39.

Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

40. 40.

Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575

(2007).

41. 41.

Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

42. 42.

Evangelou, E. & Ioannidis, J. P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).

43. 43.

Cochran, W. G. The combination of estimates from different experiments. *Biometrics* **10**, 101–129 (1954).

44. 44.

Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

45. 45.

McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

46. 46.

Kerimov, N. et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. Preprint at <https://doi.org/10.1101/2020.01.29.924266> (2021).

47. 47.

Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

48. 48.

Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

49. 49.

Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

50. 50.

Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

51. 51.

CDC. People with certain medical conditions.

<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html> (2021).

52. 52.

Williamson, E. J. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).

53. 53.

Zhou, T. et al. Educational attainment and drinking behaviors: Mendelian randomization study in UK Biobank. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-019-0596-9> (2019).

54. 54.

1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073

(2010).

55. 55.

Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human genome. *eLife* **7**, e34408 (2018).

56. 56.

Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

57. 57.

Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

58. 58.

Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the  $I^2$  statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).

59. 59.

Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44**, 313–329 (2020).

## Acknowledgements

We thank the entire COVID-19 HGI community for their contributions and continued collaboration. The work of the contributing studies was supported by numerous grants from governmental and charitable bodies.

Acknowledgements specific to contributing studies are provided in Supplementary Table 13. We thank G. Butler-Laporte, G. Wojcik, M.-G. Hollm-Delgado, C. Willer and G. Davey Smith for their extensive feedback and discussion.

# **Author information**

## Author notes

1. These authors contributed equally: Mari E. K. Niemi, Juha Karjalainen, Benjamin M. Neale, Mark Daly, Andrea Ganna
2. Lists of authors and their affiliations appear in the online version of the paper

## Affiliations

1. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

Mari E. K. Niemi, Juha Karjalainen, Mark Daly, Andrea Ganna, Juha Mehtonen, Mattia Cordioli, Mari Kaunisto, Sara Pigazzini, Kati Donner, Katja Kivinen, Aarno Palotie & Mark J. Daly

2. Broad Institute of MIT and Harvard, Cambridge, MA, USA

Rachel G. Liao, Mark Daly, Andrea Ganna, Masahiro Kanai, Kumar Veerapen, Camelia Minica, Amy Trankiem, Mary K. Balaconis, Huy Nguyen, Matthew Solomonson, Laurent Francioli, Qingbo Wang, Robert C. Green, Mark J. Daly & Sam Bryant

3. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

Mark Daly, Andrea Ganna, Hilary Finucane, Alicia R. Martin, Wei Zhou, Lindokuhle Nkambule, Konrad J. Karczewski, Elizabeth G. Atkinson, Masahiro Kanai, Kristin Tsuo, Nikolas Baya, Patrick Turley, Rahul Gupta, Raymond K. Walters, Duncan S. Palmer, Gopal Sarma, Matthew Solomonson, Nathan Cheng, Wenhan Lu, Claire Churchhouse, Jacqueline I. Goldstein, Daniel King, Cotton Seed, Mark J. Daly, Benjamin M. Neale, F. Kyle Satterstrom & Sam Bryant

4. Massachusetts General Hospital, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Benjamin M. Neale

5. Yale University, New Haven, CT, USA

Gita A. Pathak, Frank R. Wendt & Renato Polimanti

6. Icahn School of Medicine at Mount Sinai, New York, NY, USA

Shea J. Andrews, Laura Sloofman & Stuart C. Sealfon

7. Stroke Pharmacogenomics and Genetics, Biomedical Research Institute Sant Pau (IIB Sant Pau), Sant Pau Hospital, Inmungen-CoV2, Barcelona, Spain

Israel Fernandez-Cadenas

8. Institute of Virology, Technical University Munich and Helmholtz Zentrum München, Munich, Germany

Eva C. Schulte & Ulrike Protzer

9. Institute of Psychiatric Phenomics and Genomics, Medical Center of the University of Munich, Munich, Germany

Eva C. Schulte

10. Department of Psychiatry, Medical Center of the University of Munich, Munich, Germany

Eva C. Schulte

11. IRCCS, Istituto Giannina Gaslini, Genova, Italy

Pasquale Striano, Domenico A. Covello, Federico Zara, Vincenzo Salpietro, Michele Iacomino, Paolo Scudieri, Renata Bocciardi, Carlo Minetti, Maria Stella Vari & Francesca Madia

12. Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genova, Genova, Italy

Pasquale Striano

13. Queen Mary University of London, London, UK

Eirini Marouli, Stavroula Kanoni & Loukas Moutsianas

14. Open Targets, Wellcome Genome Campus, Hinxton, UK

Mohd Anisul Karim, Maya Ghoussaini, Jeremy  
Schwartzentruber & Ian Dunham

15. Department of Complex Trait Genetics, Center for Neurogenomics and  
Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit  
Amsterdam, Amsterdam, The Netherlands

Jeanne Savage, Danielle Posthuma, Elleke Tissink & Emil Uffelmann

16. Lady Davis Institute, Jewish General Hospital, McGill University,  
Montreal, Quebec, Canada

Guillaume Butler-Laporte, J. Brent Richards, Tomoko  
Nakanishi, David R. Morrison, Vincenzo Forgetta, Biswarup  
Ghosh, Laetitia Laurent, Danielle Henry, Tala Abdullah, Olumide  
Adeleye, Noor Mamlouk, Nofar Kimchi, Zaman Afrasiabi, Nardin  
Rezk, Branka Vulesevic, Meriem Bouab, Charlotte Guzman, Louis  
Petitjean, Chris Tselios, Xiaoqing Xue, Jonathan Afilalo & Darin Adra

17. Medical Research Institute, Kangbuk Samsung Hospital,  
Sungkyunkwan University School of Medicine, Suwon, Republic of  
Korea

Han-Na Kim

18. Osaka University Graduate School of Medicine, Osaka, Japan

Yukinori Okada

19. Baylor College of Medicine, Houston, TX, USA

Jinyoung Byun & Younghun Han

20. Mohammed Bin Rashid University of Medicine and Health Sciences,  
Dubai, United Arab Emirates

Mohammed Jashim Uddin

21. MRC Integrative Epidemiology Unit (IEU), University of Bristol,  
Bristol, UK

George Davey Smith

22. Department of Internal Medicine, Division of Cardiovascular  
Medicine, Michigan Medicine, Ann Arbor, MI, USA

Cristen J. Willer

23. Department of Human Genetics, University of Michigan Medical  
School, Ann Arbor, MI, USA

Cristen J. Willer

24. Department of Computational Medicine and Bioinformatics,  
University of Michigan Medical School, Ann Arbor, MI, USA

Cristen J. Willer

25. Seaver Autism Center for Research and Treatment, Department of  
Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY,  
USA

Joseph D. Buxbaum, Laura G. Sloofman, Brett L. Collins, Tess  
Levy, Slayton J. Underwood, Bari Britvan, Katherine Keller, Lara  
Tang, Michael Peruggia, Liam L. Hiester, Kristi Niblo, Alexandra  
Aksentijevich, Alexander Labkowsky, Avromie Karp & Menachem  
Zlatopolsky

26. Program in Medical and Population Genetics, Broad Institute of MIT  
and Harvard, Cambridge, MA, USA

Hilary Finucane, Alicia R. Martin, Wei Zhou, Lindokuhle  
Nkambule, Jatin Arora, Soumya Raychaudhuri, Josep  
Mercader, Konrad J. Karczewski, Elizabeth G. Atkinson, Masahiro  
Kanai, Kristin Tsuo, Nikolas Baya, Patrick Turley, Rahul  
Gupta, Raymond K. Walters, Duncan S. Palmer, Gopal  
Sarma, Matthew Solomonson, Nathan Cheng, Wenhan Lu, Claire

Churchhouse, Jacqueline I. Goldstein, Daniel King, Cotton Seed, Benjamin M. Neale, F. Kyle Satterstrom, Sam Bryant & Caroline Cusick

27. David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Bogdan Pasaniuc

28. Institut Pasteur, Paris, France

Hanna Julienne & Hugues Aschard

29. Harvard School of Public Health, Boston, MA, USA

Huwenbo Shi

30. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

Loic Yengo

31. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

Mohd Anisul Karim, Maya Ghoussaini & Jeremy Schwartzentruber

32. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK

Ian Dunham

33. Centre for Bioinformatics and Data Analysis, Medical University of Bialystok, Bialystok, Poland

Karolina Chwialkowska

34. Trieste University, Trieste, Italy

Margherita Francescatto

35. Vanderbilt University Medical Center, Nashville, TN, USA

Lea Davis & V. Eric Kerchberger

36. University of California San Francisco, San Francisco, CA, USA

Sulaggi Lee

37. Stanford University, Stanford, CA, USA

James Priest, Nirmal Vadgama & Yosuke Tanigawa

38. University of Siena, Siena, Italy

Alessandra Renieri

39. Boston Children's Hospital, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Vijay G. Sankaran

40. Blizard Institute, Queen Mary University of London, London, UK

David van Heel, Adolfo de Salazar, David A. van Heel, Karen A. Hunt & Bhavi Trivedi

41. Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands

Patrick Deelen, Lude Franke, Annique Claringbould, Esteban Lopera, Robert Warmerdam, Irene van Blokland & Pauline Lanting

42. Department of Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands

Patrick Deelen

43. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

Guillaume Butler-Laporte & J. Brent Richards

44. Department of Twin Research, King's College London, London, UK

J. Brent Richards

45. Department of Human Genetics, McGill University, Montreal, Quebec, Canada

J. Brent Richards, Tomoko Nakanishi, Vincent Mooser, Rui Li, Alexandre Belisle, Pierre Lepage, Jiannis Ragoussis, Daniel Auld & G. Mark Lathrop

46. Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Tomoko Nakanishi

47. National Institutes of Health, Bethesda, MD, USA

Les Biesecker

48. The Roslin Institute, University of Edinburgh, Edinburgh, UK

J. Kenneth Baillie, Sara Clohisey, Max Head Fourman, James Furniss, Chris Haley, Andy Law, Jonathan Millar, Erola Pairo-Castineira, Nicholas Parkinson, Konrad Rawlik, Clark D. Russell, Barbara Shih, Albert Tenesa, Bo Wang, Marie Zechner, Andrew Law, Ross Hendry, Ruth Armstrong, Ceilia Boz, Adam Brown, Louise Cullum, Nicky Day, Esther Duncan, Paul Finerman, Ailsa Golightly, Dawn Law, Rachel Law, Sarah Law, Hanning Mal, Ellie McMaster, Jen Meikle, Wilna Oosthuyzen, Trevor Paterson, Andrew Stenhouse, Maaike Swets, Helen Szoor-McElhinney, Filip Taneski, Tony Wackett, Mairi Ward, Jane Weaver, Judy Coyle, Bernadette Gallagher, Rebecca Lidstone-Scott, Debbie Hamilton & Fiona Griffiths

49. Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK

J. Kenneth Baillie, Sean Keating & Timothy Walsh

50. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK

J. Kenneth Baillie, Andrew D. Bretherick, Graeme Grimes, Caroline Hayward, Lucija Klaric, Alison M. Meynert, Erola Pairo-Castineira, Chris P. Ponting, Anne Richmond, Albert Tenesa, Veronique Vitart & James F. Wilson

51. Medical Genetics, University of Siena, Siena, Italy

Francesca Mari, Alessandra Renieri, Mirjam Lista, Valentina Perticaroli, Chiara Fallerini, Sergio Daga, Margherita Baldassarri, Francesca Fava, Elisa Frullanti, Floriana Valentino, Gabriella Doddato, Annarita Giliberti, Mirella Bruttini, Susanna Croci, Ilaria Meloni, Giada Beligni, Andrea Tommasi, Laura Di Sarno, Maria Palmieri, Miriam Lucia Carriero, Diana Alaverdian, Nicola Iuso & Gabriele Inchingolo

52. Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy

Francesca Mari, Alessandra Renieri, Valentina Perticaroli, Francesca Fava, Rossella Tita, Sara Amitrano, Mirella Bruttini, Maria Antonietta Mencarelli, Caterina Lo Rizzo, Anna Maria Pinto & Andrea Tommasi

53. Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Siena, Italy

Alessandra Renieri, Francesca Mari, Alessandra Renieri, Francesca Montagnani, Mario Tumbarello, Mirjam Lista, Valentina Perticaroli, Simone Furini, Elisa Benetti, Kristina Zguro, Katia Capitani, Chiara Fallerini, Sergio Daga, Margherita Baldassarri, Francesca Fava, Elisa Frullanti, Floriana Valentino, Gabriella Doddato, Annarita Giliberti, Mirella Bruttini, Susanna Croci, Ilaria Meloni, Giada Beligni, Andrea

Tommasi, Laura Di Sarno, Maria Palmieri, Miriam Lucia  
Carrieri, Diana Alaverdian, Nicola Iuso, Gabriele  
Inchingolo, Giacomo Zanelli & Francesco Bianchi

54. Department of Electronics, Information and Bioengineering (DEIB),  
Politecnico di Milano, Milano, Italy

Anna Bernasconi, Stefano Ceri & Pietro Pinoli

55. Politecnico di Milano, Milan, Italy

Stefano Ceri Baillie & Arif Canakoglu

56. University of Michigan, Ann Arbor, MI, USA

Brooke Wolford

57. Vanderbilt School of Medicine, Nashville, TN, USA

Annika Faucon

58. All India Institute of Medical Sciences Kalyani, Kalyani, India

Atanu Kumar Dutta & Sanjay Patel

59. Hasso Plattner Institute, New York, NY, USA

Claudia Schurmann

60. Naina Tech, Hyderabad, India

Emi Harry

61. EMBL-European Bioinformatics Institute, Hinxton, UK

Ewan Birney

62. University of Northampton, Northampton, UK

Jamal Nasir

63. University of Helsinki, Helsinki, Finland

Minttu Marttila

64. University of Miami, Miami, FL, USA

Nicole Dueker

65. Ecole Centrale de Nantes, Inserm, Centre de Recherche en  
Transplantation et Immunologie, Nantes University, UMR1064, ITUN,  
Nantes, France

Sophie Limou

66. University of Liège, Liège, Belgium

Souad Rahmouni

67. Qatar Genome Program, Qatar Foundation Research, Development  
and Innovation, Qatar Foundation, Doha, Qatar

Hamdi Mbarek, Dima Darwish, Said I. Ismail, Chadi Saad, Yaser Al-  
Sarraj, Radja Messai Badji, Wadha Al-Muftah & Asma Al Thani

68. Medical and Population Genetics and Cardiovascular Disease  
Initiative, Broad Institute of Harvard and MIT, Cambridge, Cambridge,  
MA, USA

Md Mesbah Uddin

69. Cardiovascular Research Center, Massachusetts General Hospital,  
Boston, MA, USA

Md Mesbah Uddin

70. Intensive Care Unit, Vall d'Hebron Hospital, Barcelona, Spain

Raquel Albertos & Ricard Ferrer

71. Institut de Biomedicina de València - CSIC, València, Spain

Jordi Pérez-Tur

72. Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), València, Spain

Jordi Pérez-Tur

73. Unidad Mixta de Neurología y Genética, Instituto de Investigación Sanitaria La Fe, València, Spain

Jordi Pérez-Tur

74. Erasmus Medical Center, Rotterdam, The Netherlands

Ruolin Li, Carolina Medina-Gómez, Bahar Sedaghati-Khayat, Costanza Vallerga, Cindy Boer, Ingrid Kennis-Szilagy, Vid Prijatelj & Ivana Prokić

75. National Genome Center, Copenhagen, Denmark

Lasse Folkersen

76. University of Copenhagen, Copenhagen, Denmark

Ida Moltke

77. Genomics PLC, Oxford, UK

Nils Koelling & Chris A. Spencer

78. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

Alexander Teumer

79. Department of Population Medicine and Lifestyle Diseases Prevention,  
Medical University of Białystok, Białystok, Poland

Alexander Teumer

80. Genomics England, London, UK

Athanasiou Kousathanas, Dorota Pasko, Loukas Moutsianas, Mark J. Caulfield, Richard H. Scott, Susan Walker, Alex Stuckey, Christopher A. Odhams, Daniel Rhodes, Tom Fowler, Augusto Rendon, Georgia Chan, Prabhu Arumugam & Richard H. Scott

81. Junta de Andalucía, Seville, Spain

Alicia Utrilla

82. Human Genetics Program of ICBM and Department of Basic-Clinical Oncology, University of Chile, Santiago, Chile

Ricardo A. Verdugo

83. Center for the Development of Scientific Research (CEDIC),  
Asunción, Paraguay

Ruth Zárate

84. Translational Bioinformatics Unit, Navarrabiomed, Complejo Hospitalario de Navarra (CHN), Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain

David Gómez-Cabrero

85. Mucosal & Salivary Biology Division, King's College London Dental Institute, London, UK

David Gómez-Cabrero

86. GENYO, Center for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, Granada,

Spain

Elena Carnero-Montoro, Marta E. Alarcón-Riquelme & Manuel Martínez-Bueno

87. University of Puerto Rico, San Juan, Puerto Rico

Carmen L. Cadilla

88. National Laboratory of Genomics for Biodiversity (LANGEBIO), Advanced Genomics Unit, CINVESTAV, Irapuato, Mexico

Andrés Moreno-Estrada & Adriana Garmendia

89. Queensland University of Technology, Brisbane, Queensland, Australia

Leire Moya

90. Clinical Research Unit of Nanoro, Institut de Recherche en Sciences de la Santé, CNRST, Ouagadougou, Burkina Faso

Palwendé Romuald Boua

91. McGill University, Montreal, Quebec, Canada

Marie-Julie Favé

92. Université de Montréal, Montreal, Quebec, Canada

Audrey Lemaçon

93. Fonds de la Recherche Scientifique (FNRS) & Centre de Génétique Humaine, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

Isabelle Migeotte

94. University of Pecs Medical School, Pécs, Hungary

Reka Varnai & Jozsef L. Szentpeteri

95. Institute of Biomedicine and Cancer Research Laboratories, Western Cancer Centre FICAN West, University of Turku, Turku, Finland

Csilla Sipeky

96. Institute of Biomedical Technologies, National Research Council, Segrate, Italy

Francesca Colombo

97. Immediate, Milan, Italy

Kathrin von Hohenstaufen

98. University of Cambridge, Cambridge, UK

Pietro Lio

99. Genome Opinion, Seoul, Republic of Korea

Hogune Im, Chulho Han, Han Song, Jiwoo Lim, Younhe Lee & Sugyeong Kim

100. University of Groningen, Groningen, The Netherlands

Biljana Atanasovska

101. Universiti Malaysia Pahang, Gambang, Malaysia

Hajar Fauzan Ahmad

102. Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Philip Jansen, Jouke Jan Hottenga, Meike Bartels, Eco J. C. de Geus & Michel G. Nivard

103. University Medical Centre Groningen, University of Groningen,  
Groningen, The Netherlands

Lude Franke

104. MNM DIAGNOSTICS, Pozna?, Poland

Elżbieta Kaja

105. Institute for Systems Biology, Seattle, WA, USA

Sergey A. Kornilov

106. Sultan Idris Education University, Tanjung Malim, Malaysia

Ilangkumaran Sivanadhan

107. Hospital Kulim, Kedah, Malaysia

Sarala Perumal

108. AbbVie, Lake Buff, IL, USA

Sahar Esmaeeli

109. Root Deep Insight, Boston MA, USA

Nathaniel M. Pearson

110. 23andMe, Sunnyvale, CA, USA

Adam Auton, Janie F. Shelton, Anjali J. Shastri, Teresa Filshtein-Sonmez, Daniella Coker, Antony Symons, Stella Aslibekyan, Jared O'Connell, Chelsea Ye & Catherine H. Weldon

111. GSK, Stevenage, UK

Jorge Esparza-Gordillo

112. Department of Pharmacology, Feinberg School of Medicine,  
Northwestern University, Chicago, IL, USA

Minoli Perera, Guang Yang, Cristina Alarcon, Stefanie  
Herrmann & Paula Friedman

113. Department of Medicine, Northwestern University, Chicago, IL, USA

Kevin O'Leary & Sophia Mazurek

114. Washington DC Veterans Affairs Medical Center, Hospital Medicine,  
Washington, DC, USA

Matthew Tuck, Jeff Banagan & Zacharia Hamidi

115. Department of Medicine, George Washington University, Washington,  
DC, USA

Travis O'Brien

116. Section of Hospital Medicine, Department of Medicine, University of  
Chicago, Chicago, IL, USA

David Meltzer, Noora Raffat & Carmen de la Horra

117. Section of Hematology and Oncology, Department of Medicine,  
University of Chicago, Chicago, IL, USA

Peter O'Donnell

118. College of Pharmacy, University of Illinois at Chicago, Chicago, IL,  
USA

Edith Nutescu & Diana Moreno

119. Department of Pharmacology, George Washington University,  
Washington, DC, USA

April Barbour & Rocío Gallego-Durán

120. Department of Neurology, Amsterdam UMC, Amsterdam  
Neuroscience, Amsterdam, The Netherlands

Bart Ferwerda, Diederik van de Beek, Matthijs C. Brouwer, Martijn  
Beudel & Rutger Koning

121. Department of Intensive Care, Amsterdam UMC, Amsterdam, The  
Netherlands

Alexander P. J. Vlaar, Anne Geke Algera, Frank van Baarle, Lieuwe  
Bos, Michela Botta, Sanne de Bruin, Esther Bulle, Dave  
Dongelmans, Paul Elbers, Lucas Fleuren, Armand Girbes, Laura  
Hagens, Leo Heunks, Janneke Horn, Niels van Mourik, Frederique  
Paulus, Jorinde Raasveld, Marcus J. Schultz, Marry Smit, Charlotte  
Teunissen, Patrick Thoral & Heder de Vries

122. Department of Infectious Diseases, Amsterdam UMC, Amsterdam,  
The Netherlands

W. Joost Wiersinga, Michiel van Agtmael, Marije Bomers, Justin de  
Brabander, Godelieve de Bree, Osoul Chouchane, Suzanne  
Geerlings, Bram Goorhuis, Martin P. Grobusch, Vanessa  
Harris, Sabine M. Hermans, Joppe W. Hovius, Jeannine Nellen, Edgar  
Peters, Tom van der Poll, Jan M. Prins, Tom Reijnders, Michiel  
Schinkel, Alex Schuurman, Kim Sigaloff, Cornelis S. Stijnis, Anissa  
Tsonas, Marc van der Valk & Michèle van Vugt

123. Department of Clinical Epidemiology, Biostatistics and  
Bioinformatics, Amsterdam UMC, Amsterdam, The Netherlands

A. H. Koos Zwinderman

124. Experimental Immunology, Amsterdam UMC, Amsterdam, The  
Netherlands

Diane Bax, Alex Cloherty, Theo Geijtenbeek, Florianne  
Hafkamp & Bennedikt Preckel

125. Department of Pulmonology, Amsterdam UMC, Amsterdam, The Netherlands

Harm Jan Bogaard, Peter I. Bonta & Esther J. Nossent

126. Department of Pathology, Amsterdam UMC, Amsterdam, The Netherlands

Marianna Bugiani

127. Department of Anesthesiology, Amsterdam UMC, Amsterdam, The Netherlands

Bart Geerts, Markus Hollmann & Denise Veelo

128. Amsterdam UMC Biobank Core Facility, Amsterdam UMC, Amsterdam, The Netherlands

Jorg Hamann

129. Department of Radiology, Amsterdam UMC, Amsterdam, The Netherlands

Robert Hemke

130. Department of Medical Microbiology, Amsterdam UMC, Amsterdam, The Netherlands

Menno D. de Jong

131. Department of Clinical Chemistry, Amsterdam UMC, Amsterdam, The Netherlands

Willemke Stilma & Dorien Wouters

132. Amsterdam UMC Biobank, Amsterdam UMC, Amsterdam, The Netherlands

René P. Minnaar, Adrie Kromhout & Kees W. J. van Uffelen

133. Core Facility Genomics, Amsterdam UMC, Amsterdam, The Netherlands

Ruud A. Wolterman

134. Ancestry, Lehi, UT, USA

Genevieve Roberts, Danny Park, Catherine A. Ball, Marie Coignet, Shannon McCurdy, Spencer Knight, Raghavendran Partha, Brooke Rhead, Miao Zhang, Nathan Berkowitz, Michael Gaddis, Keith Noto, Luong Ruiz, Milos Pavlovic, Eurie L. Hong, Kristin Rand, Ahna Girshick, Harendra Guturu & Asher Haug Baltzell

135. GIGA-Institute, University of Liège, Liège, Belgium

Souad Rahmouni, Michel Georges, Sofia Melo, Nicolas Jacques, Emmanuel Di Valentin, François Giroule, Alice Collignon, Coraline Radermecker, Marielle Lebrun, Hélène Perée, Samuel Latour, Olivia Barada, Judit Sanchez, Margot Meunier, Emeline Mariavelle, Sandy Anania, Hélène Gazon, Myriam Mni, Marie Wéry & Yasmine Belhaj

136. CHC Mont-Légia, Liège, Belgium

Julien Guntz, Laurent Jadot & Sabine Claassen

137. BHUL (Liège Biobank), CHU of Liège, Liège, Belgium

Yves Beguin, Stéphanie Gofflot, Kamilia El Kandoussi & Raphaël Thonon

138. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Lindokuhle Nkambule

139. Centre de Génétique Humaine, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

Youssef Bouysran, Adeline Busson, Xavier Peyrassol, Françoise Wilkin, Bruno Pichon, Guillaume Smits & Isabelle Vandernoot

140. Service de Médecine Interne, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

Jean-Christophe Goffard

141. CHU of Liège, University of Liège, Liège, Belgium

Michel Moutschen, Benoit Missé, Gilles Darcis, Julien Guiot, Samira Azarzar, Patricia Dellot, Axelle Bertrand, Gilles Parzibut, Mathilde Clarinval, Catherine Moermans, Olivier Malaise, Pascale Huynen, Alyssia Mesdagh, Claire Josse, Bouchra Boujemla, Danusia Juszczak, Marjorie Fadur, Séverine Camby, Christelle Meuris, Marie Thys, Jessica Jacques, Monique Henket, Philippe Léonard, Frédéric Frippiat, Jean-Baptiste Giot, Anne-Sophie Sauvage, Christian Von Frenckell, Alicia Staderoli, Bernard Lamberton & Edouard Louis

142. McGill Genome Centre, McGill University, Montréal, Québec, Canada

Vincent Mooser, Rui Li, Alexandre Belisle, Pierre Lepage, Jiannis Ragoussis, Daniel Auld & G. Mark Lathrop

143. Department of Emergency Medicine, McGill University, Montreal, Quebec, Canada

Marc Afilalo

144. Emergency Department, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

Marc Afilalo

145. McGill AIDS Centre, Department of Microbiology and Immunology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

Maureen Oliveira

146. McGill Centre for Viral Diseases, Department of Infectious Disease,  
Lady Davis Institute, Jewish General Hospital, Montreal, Quebec,  
Canada

Bluma Brenner

147. Research Centre of the Centre Hospitalier de l'Université de Montréal,  
Montreal, Quebec, Canada

Nathalie Brassard

148. Department of Medicine, Research Centre of the Centre Hospitalier de  
l'Université de Montréal, Montreal, Quebec, Canada

Madeleine Durand

149. Department of Medicine, Université de Montréal, Montreal, Quebec,  
Canada

Madeleine Durand, Michaël Chassé & Daniel E. Kaufmann

150. Department of Medicine and Human Genetics, McGill University,  
Montreal, Quebec, Canada

Erwin Schurr

151. Department of Intensive Care, Research Centre of the Centre  
Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada

Michaël Chassé

152. Division of Infectious Diseases, Research Centre of the Centre  
Hospitalier de l'Université de Montréal, Montréal, Quebec, Canada

Daniel E. Kaufmann

153. Division of Genetic Medicine, Department of Medicine, Vanderbilt  
University Medical Center, Nashville, TN, USA

Lea K. Davis, Nancy J. Cox, Jennifer E. Below, Julia M. Sealock, Annika B. Faucon, Megan M. Shuey, Hannah G. Polikowsky, Lauren E. Petty, Douglas M. Shaw, Hung-Hsin Chen & Wanying Zhu

154. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

Lea K. Davis, Nancy J. Cox, Jennifer E. Below, Julia M. Sealock, Annika B. Faucon, Megan M. Shuey, Hannah G. Polikowsky, Lauren E. Petty, Douglas M. Shaw, Hung-Hsin Chen & Wanying Zhu

155. Institute of Human Genetics, University Hospital Bonn, Medical Faculty University of Bonn, Bonn, Germany

Kerstin U. Ludwig, Julia Schröder, Selina Rolker, Markus M. Nöthen & Julia Fazaal

156. Institute of Genomic Statistics and Bioinformatics, University Hospital Bonn, Medical Faculty University of Bonn, Bonn, Germany

Carlo Maj

157. Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Medical Faculty Heinrich Heine University, Düsseldorf, Germany

Verena Keitel, Björn-Erik Ole Jensen, Torsten Feldt & Lisa Knopp

158. Institute of Human Genetics, Medical Faculty, RWTH Aachen University, Aachen, Germany

Ingo Kurth & Thomas Eggemann

159. Clinic for Cardiology, Angiology and Internal Intensive Medicine, Medical Clinic I, RWTH Aachen University, Aachen, Germany

Nikolaus Marx

160. Department of Pneumology and Intensive Care Medicine, Faculty of Medicine, RWTH Aachen University, Aachen, Germany

Michael Dreher

161. Department of Pneumology, Hannover Medical School, Hannover, Germany

Isabell Pink

162. Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

Markus Cornberg

163. Hannover Unified Biobank, Hannover Medical School, Hannover, Germany

Thomas Illig & Sonja Volland

164. Department I of Internal Medicine, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Cologne, Germany

Clara Lehmann, Philipp Schommers, Max Augustin, Jan Rybniker, Philipp Koehler, Jan Rybniker & Oliver A. Cornely

165. Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany

Clara Lehmann & Philipp Schommers

166. German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany

Clara Lehmann & Philipp Schommers

167. Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany

Janine Altmüller

168. Department of Anesthesiology and Intensive Care Medicine,  
University Hospital Essen, University Duisburg-Essen, Essen,  
Germany

Marc M. Berger & Thorsten Brenner

169. Department of Child and Adolescent Psychiatry, University Hospital  
Essen, University of Duisburg-Essen, Essen, Germany

Anke Hinney

170. Department of Infectious Diseases, University Hospital Essen,  
University Duisburg-Essen, Essen, Germany

Oliver Witzke

171. Department of Pneumology, Allergology and Respiratory Medicine,  
University Hospital Saarland, Homburg/Saar, Germany

Robert Bals & Christian Herr

172. Center of Human and Molecular Biology, Department of Human  
Genetics, University Hospital Saarland, Homburg/Saar, Germany

Nicole Ludwig

173. Department of Genetics & Epigenetics, Saarland University,  
Saarbrücken, Germany

Jörn Walter

174. Eurac Research, Institute for Biomedicine (affiliated to the University  
of Lübeck), Bolzano, Italy

Christian Fuchsberger, Cristian Pattaro, Alessandro De Grandi, Peter  
Pramstaller, David Emmert, Roberto Melotti, Luisa Foco, Deborah  
Mascalzoni, Martin Gögele, Francisco Domingues & Andrew Hicks

175. University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Christopher R. Gignoux, Stephen J. Wicks, Kristy Crooks, Kathleen C. Barnes, Michelle Daya, Jonathan Shortt, Nicholas Rafaels & Sameer Chavan

176. Department of Genetics and Development, Institute for Genomic Medicine, Columbia University, New York, NY, USA

David B. Goldstein

177. Department of Medicine, Institute for Genomic Medicine, Columbia University, New York, NY, USA

Krzysztof Kiryluk

178. Department of Biomedical Informatics, Columbia University, New York, NY, USA

Soumitra Sengupta, Amritha Menon & Yat S. So

179. Department of Pediatrics, Columbia University, New York, NY, USA

Wendy Chung

180. Department of Medicine, Columbia University, New York, NY, USA

Muredach P. Reilly, Atlas Khan, Chen Wang, Ali G. Gharavi, Ning Shang, Sheila M. O'Byrne & Renu Nandakumar

181. Institute for Genomic Medicine, Columbia University, New York, NY, USA

Gundula Povysil & Nitin Bhardwaj

182. Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA

Iuliana Ionita-Laza

183. Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

Eldad Hod & Danielle Pendrick

184. Medical Research Institute, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Han-Na Kim

185. Department of Clinical Research Design and Evaluation, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

Han-Na Kim

186. Division of Gastroenterology, Department of Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University, School of Medicine, Seoul, Republic of Korea

Soo-Kyung Park

187. Department of Biochemistry, College of Medicine, Ewha Womans University, Seoul, Republic of Korea

Hyung-Lae Kim

188. Department of Internal Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

Chang Kyung Kang & Wan Beom Park

189. Department of Periodontology, Section of Dentistry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Hyo-Jung Lee

190. Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea

Kyoung-Ho Song, Sin Young Ham, Jongtak Jung, Eu Suk Kim & Hong Bin Kim

191. Department of Physical & Rehabilitation Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Kyung Jae Yoon

192. Department of Clinical Research Design & Evaluation, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

Kyung Jae Yoon

193. Biomedical Institute for Convergence at SKKU, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea

Kyung Jae Yoon

194. Department of Public Health Service, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Nam-Jong Paik

195. Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

Nam-Jong Paik

196. Korea Research Environment Open NETwork, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea

Woojin Seok

197. Global Science Experimental Data Hub Center, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea

Heejun Yoon

198. Division of Infectious Diseases, Department of Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Eun-Jeong Joo

199. Center for Cohort Studies, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Yoosoo Chang & Seungho Ryu

200. Department of Occupational and Environmental Medicine, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Yoosoo Chang & Seungho Ryu

201. Department of Laboratory Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea

Jeong Su Park & Kyoung Un Park

202. Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany

David Ellinghaus, Frauke Degenhardt, Simonas Juzenas, Florian Tran, Philip Rosenstiel, Stefan Schreiber, Andre Franke, Eike Matthias Wacker, Florian Uellendahl-Werth, Hesham ElAbd, Lars Wienbrandt, Malte Christoph Rühlemann, Mareike Wendorff, Anja Tanck, Christoph Gassner, Georg Hemmrich-Stanisak, Jan Kässens, Maria E. Figuera Basso, Martin Schulzky, Michael

Wittig, Nicole Braun, Tanja Wesse, Wolfgang Albrecht, Xiaoli Yi, Almut Nebel, Anna Peschuck & Sandra May

203. Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

David Ellinghaus, Banasik Karina & Søren Brunak

204. Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain

Mario Cáceres & Jon Lerga-Jaso

205. ICREA, Barcelona, Spain

Mario Cáceres

206. Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

Tobias L. Lenz, Ana Teles, Clinton Azuure & Onur Özer

207. Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, Hamburg, Germany

Tobias L. Lenz, Ana Teles & Onur Özer

208. Department of Gastroenterology, Hospital Universitario Ramón y Cajal, University of Alcalá, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain

Agustín Albillos, Beatriz Mateos, Francisco Mesonero, Luis Téllez & Miguel Rodríguez-Gandía

209. Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

Agustín Albillos, Luis Bujanda, Jesus M. Banales, Manuel Romero-Gómez, Maria Buti, Douglas Maya-Miles, Beatriz Mateos, Francisco Mesonero, Koldo Garcia-Etxebarria, Laura Izquierdo-Sánchez, Luis Téllez, Miguel Rodríguez-Gandía, Pedro M. Rodrigues, Francisco Rodriguez-Frias, Mar Riveiro-Barciela & Rocío Gallego-Durán

210. Vall d' Hebron Institut de Recerca (VHIR), Vall d' Hebron Hospital Universitari, Barcelona, Spain

Antonio Julià, Adriana Palom, Francisco Rodriguez-Frias & Sara Marsal

211. Charite Universitätsmedizin Berlin, Berlin, Germany

Bettina Heidecker, Florian Kurth, Leif Erik Sander, Alena Mayer, Alice Braun, Carsten Skurk, Charlotte Thibeault, Elisa T. Helbig, Julia Kraft, Lena J. Lippert, Phillip Suwalski, Stephan Ripke, Wolfgang Poller, Xiaomin Wang & Zehra Karadeniz

212. Hospital Universitario Clinico San Cecilio, Granada, Spain

Federico Garcia, Adolfo de Salazar, Jose Hernández Quero, Natalia Chueca & Trinidad Gonzalez Cejudo

213. Instituto de Investigación Ibs.Granada, Granada, Spain

Federico Garcia & Jose Hernández Quero

214. Emergency Department, University Hospital Regensburg, Regensburg, Germany

Frank Hanses

215. Department for Infectious Diseases and Infection Control, University Hospital Regensburg, Regensburg, Germany

Frank Hanses & Sibylle Wilfling

216. Medical University of Innsbruck, Department of Medicine and Christian Doppler Laboratory on Iron and Phosphate Biology, Innsbruck, Austria

Heinz Zoller & Benedikt Schaefer

217. Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Jan C. Holter, Johannes R. Hov, Tom H. Karlsen, Trine Folseraas, Aleksander Rygh Holten, Anne Ma Dyrhol-Riise, Kristian Tonby, Andreas Lind, Fredrik Müller, Susanne Dudman & Marit M. Grimsrud

218. Department of Microbiology, Oslo University Hospital, Oslo, Norway

Jan C. Holter, Andreas Lind, Fredrik Müller & Susanne Dudman

219. Hospital Clinic, University of Barcelona and IDIBAPS, Barcelona, Spain

Javier Fernández, Carlos Ferrando, Enric Reverter, Joan Ramon Badia, María Hernandez-Tejero & Pedro Castro

220. European Foundation for the Study of Chronic Liver Failure (EF-CLIF), Barcelona, Spain

Javier Fernández

221. Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany

Philipp Koehler & Oliver A. Cornely

222. Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany

Philipp Koehler & Jan Rybniker

223. Genomes for Life-GCAT labGermans Trias i Pujol Research Institute (IGTP), Badalona, Spain

Rafael de Cid, Anna Carreras Nolla, Beatriz Cortes, Iván Galván-Femenia, Natalia Blay & Xavier Farre

224. IRCCS Humanitas Research Hospital, Milan, Italy

Rosanna Asselta, Stefano Duga, Alberto Mantovani, Alessandro Protti, Alessio Aghemo, Ana Lleo, Antonio Voza, Elena Azzolini, Elvezia Maria Paraboschi, Ilaria My, Maurizio Cecconi, Paolo Tentorio, Salvatore Badalamenti, Sara Bombace & Stefano Duga

225. Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

Rosanna Asselta, Stefano Duga, Alberto Mantovani, Alessandro Protti, Alessio Aghemo, Ana Lleo, Claudio Cappadona, Elena Azzolini, Elvezia Maria Paraboschi, Maurizio Cecconi, Valeria Rimoldi & Stefano Duga

226. Institute of Transfusionsmedicine, University Hospital Schleswig-Holstein (UKSH), Kiel, Germany

Siegfried Goerg

227. Klinik für Innere Medizin I, Universitätsklinikum Schleswig-Holstein, Kiel Campus, Kiel, Germany

Stefan Schreiber

228. Zentrum für Humangenetik Regensburg, Regensburg, Germany

Ute Hehr & Sibylle Wilfling

229. University Hospital Schleswig-Holstein (UKSH), Kiel Campus, Kiel, Germany

Andre Franke & Nicole Braun

230. Section for Gastroenterology, Department of Transplantation Medicine, Division for Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway

Johannes R. Hov, Tom H. Karlsen & Trine Folseraas

231. Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway

Johannes R. Hov, Tom H. Karlsen, Trine Folseraas & Marit M. Grimsrud

232. Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway

Johannes R. Hov, Tom H. Karlsen, Trine Folseraas & Marit M. Grimsrud

233. Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Jatin Arora & Soumya Raychaudhuri

234. Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

Jatin Arora, Soumya Raychaudhuri, Robert C. Green & Emma F. Perez

235. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

Jatin Arora, Soumya Raychaudhuri & Masahiro Kanai

236. Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA

Jatin Arora & Soumya Raychaudhuri

237. Randaberg Municipality, Randaberg, Norway

Ole Bernt Lenning

238. Department of Quality and Health Technology, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

May Sissel Vadla

239. Department of Genetics and Bioinformatics (HDGB), Division of Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway

Ronny Myhre

240. Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK

Soumya Raychaudhuri

241. Department of Intensive Care, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain

Aaron Blandino Ortiz & Raúl de Pablo

242. Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Clinical Biochemistry Department, San Sebastian, Spain

Adolfo Garrido Chercoles & Beatriz Nafria-Jimenez

243. Research Center and Memory Clinic, Fundació ACE, Institut Català de Neurociències Aplicades, Universitat Internacional de Catalunya,

Barcelona, Spain

Agustín Ruiz, Isabel Hernández, Itziar de Rojas, Marta  
Marquié & Mercè Boada

244. Networking Research Center on Neurodegenerative Diseases  
(CIBERNED), Instituto de Salud Carlos III, Madrid, Spain

Agustín Ruiz, Isabel Hernández, Itziar de Rojas, Marta  
Marquié & Mercè Boada

245. Department of Acute Medicine, Oslo University Hospital, Oslo,  
Norway

Aleksander Rygh Holten

246. Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan,  
Italy

Daniele Prati, Guido Baselli, Luca Valenti, Alessandro  
Cherubini, Antonio Muscatello, Cinzia Hu, Cristiana Bianco, Daniela  
Galimberti, Elio Scarpini, Ferruccio Ceriotti, Giuseppe  
Lamorte, Leonardo Terranova, Mahnoosh Ostadreza, Alberto  
Zanella, Alessandra Bandera, Andrea Gori, Anna Ludovica  
Fracanzani, Antonio Pesenti, Cinzia Paccapelo, Filippo Martinelli-  
Boneschi, Flora Peyvandi, Francesco Blasi, Giacomo  
Grasselli, Giorgio Costantino, Luigi Santoro, Luigia Scudeller, Maria  
Carrabba, Marina Baldini, Monica Miozzo, Nicola Montano, Roberta  
Gualtierotti, Serena Pelusi, Silvano Bosari, Stefano Aliberti & Valter  
Monzani

247. European Reference Network on Hepatological Diseases (ERN RARE  
LIVER), San Gerardo Hospital, Monza, Italy

Pietro Invernizzi, Alessio Gerussi & Chiara Milani

248. Division of Gastroenterology, Center for Autoimmune Liver Diseases,  
Department of Medicine and Surgery, University of Milan Bicocca,

Milan, Italy

Pietro Invernizzi, Alessio Gerussi & Chiara Milani

249. German Center for Neurodegenerative Diseases (DZNE Bonn), Bonn, Germany

Alfredo Ramirez

250. Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Medical Faculty, University of Cologne, Cologne, Germany

Alfredo Ramirez & Victor Andrade

251. Department of Psychiatry, Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA

Alfredo Ramirez

252. Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany

Alfredo Ramirez & Victor Andrade

253. Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain

Maria Buti, Ana Barreira, Adriana Palom & Mar Riveiro-Barciela

254. Department of Anesthesiology and Intensive Care, University Hospital of North Norway, Tromsø, Norway

Anders Benjamin Kildal

255. Klinik für Innere Medizin I, Universitätsklinikum Schleswig-Holstein, Kiel Campus, Kiel, Germany

Andreas Glück & Thomas Bahmer

256. Gastroenterology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

Anna Latiano & Orazio Palmieri

257. Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway

Anne Ma Dyrhol-Riise

258. Microbiology Department, Hospital Universitari Vall d' Hebron, Barcelona, Spain

Ariadna Rando-Segura

259. Universitat Autònoma de Barcelona, Bellaterra, Spain

Maria Buti, Ariadna Rando-Segura, Luisa Roade, Francisco Rodriguez-Frias & Mar Riveiro-Barciela

260. Department of Respiratory Diseases, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain

Aurora Solier, David Jiménez & Rosa Nieto

261. Department of Respiratory Medicine and Allergology, University Hospital, Goethe University, Frankfurt am Main, Germany

Carla Bellinghausen

262. Department of Infectious Diseases, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain

Carmen Quereda & Enrique Navas

263. Department of Internal Medicine II, Technical University of Munich,  
School of Medicine, University Hospital rechts der Isar, Munich,  
Germany

Christoph D. Spinner & Jochen Schneider

264. Division of Clinical Infectious Diseases, Research Center Borstel,  
Borstel, Germany

Christoph Lange & Jan Heyckendorf

265. German Center for Infection Research (DZIF) Clinical Tuberculosis  
Unit, Borstel, Germany

Christoph Lange & Jan Heyckendorf

266. Respiratory Medicine & International Health, University of Lübeck,  
Lübeck, Germany

Christoph Lange & Jan Heyckendorf

267. Osakidetza Basque Health Service, Basurto University Hospital,  
Respiratory Service, Bilbao, Spain

Cristina Sancho & Maider Intxausti

268. Department of Clinical and Molecular Medicine, Faculty of Medicine  
and Health Science, Norwegian University of Science and Technology,  
Trondheim, Norway

Dag Arne Lihaug Hoff & Jan Egil Afset

269. Clinic Ålesund Hospital, Department of Medicine, Møre & Romsdal  
Hospital Trust, Ålesund, Norway

Dag Arne Lihaug Hoff & Sammra Haider

270. Department of Anesthesiology, Hospital Universitario Ramón y Cajal,  
Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS),

University of Alcalá, Madrid, Spain

David Pestaña

271. Spain Hospital Clinic, University of Barcelona and IDIBAPS,  
Barcelona, Spain

David Toapanta

272. Osakidetza Basque Health Service, Galdakao Hospital, Respiratory  
Service, Galdakao, Spain

Eloisa Urrechaga & Pedro P. España

273. IBMDR - E.OOspedali Galliera, Genova, Italy

Emanuele Pontali, Lucia Garbarino & Michela Mazzocco

274. Liver ICU, Hospital Clinic Barcelona, Barcelona, Spain

David Toapanta & Enric Reverter

275. Biocruces Bizkaia Health Research Institute, Barakaldo, Spain

Eunate Arana & Natale Imaz Ayo

276. Histocompatibilidad y Biología Molecular, Centro de Transfusión de  
Madrid, Madrid, Spain

Félix García Sánchez

277. University of Milan, Milan, Italy

Luca Valenti, Francesco Malvestiti, Alberto Zanella, Alessandra  
Bandera, Andrea Gori, Anna Ludovica Fracanzani, Antonio  
Pesenti, Filippo Martinelli-Boneschi, Flora Peyvandi, Giacomo  
Grasselli, Giorgio Costantino, Monica Miozzo, Nicola  
Montano, Roberta Gualtierotti, Serena Pelusi & Silvano Bosari

278. Fondazione Grigioni per il Morbo di Parkinson, Milan, Italy

Gianni Pezzoli

279. Department of Anesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Frankfurt, Frankfurt am Main, Germany

Holger Neb

280. German Center for Infection Research (DZIF), Medical Faculty and University Hospital Cologne, University of Cologne, Partner Site Bonn-Cologne, Cologne, Germany

Jan Rybniker & Oliver A. Cornely

281. Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Jesus M. Banales & Mauro D' Amato

282. Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute, Donostia University Hospital, University of the Basque Country (UPV/EHU), San Sebastian, Spain

Luis Bujanda, Jesus M. Banales, Mauro D' Amato, Koldo Garcia-Etxebarria, Laura Izquierdo-Sanchez & Pedro M. Rodrigues

283. Infectious Diseases Service, Osakidetza, Biocruces Bizkaia Health Research Institute, Barakaldo, Spain

Josune Goikoetxea

284. Medical Department, Drammen Hospital, Vestre Viken Hospital Trust, Drammen, Norway

Karl Erik Müller & Lars Heggelund

285. Research Center Borstel, BioMaterialBank Nord, Borstel, Germany

Karoline I. Gaede

286. German Center for Lung Research (DZL), Airway Research Center North (ARCN), Giessen, Germany

Karoline I. Gaede

287. Popgen 2.0 Network (P2N), Kiel, Germany

Karoline I. Gaede

288. Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute, Donostia University Hospital, University of the Basque Country (UPV/EHU), CIBERehd, San Sebastian, Spain

Luis Bujanda & Koldo Garcia-Etxebarria

289. Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway

Kristian Tonby

290. Department of Clinical Science, University of Bergen, Bergen, Norway

Lars Heggelund

291. Biodonostia Health Research Institute, Donostia University Hospital, San Sebastian, Spain

Pedro M. Rodrigues

292. Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain

Lauro Sumoy

293. ISGlobal, Barcelona, Spain

Manolis Kogevinas, Judith Garcia-Aymerich, Gemma Castaño-Vinyals & Carlota Dobaño

294. CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

Manolis Kogevinas, Victor Moreno, Carmen de la Horra, Judith Garcia-Aymerich, Gemma Castaño-Vinyals & Carlota Dobaño

295. Universitat Pompeu Fabra (UPF), Barcelona, Spain

Manolis Kogevinas, Judith Garcia-Aymerich & Gemma Castaño-Vinyals

296. Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain

Manolis Kogevinas & Gemma Castaño-Vinyals

297. Osakidetza Basque Health Service, Donostialdea Integrated Health Organization, San Sebastian, Spain

María A. Gutiérrez-Stampa

298. Department of Internal Medicine, Infectious Diseases, University Hospital Frankfurt and Goethe University Frankfurt, Frankfurt am Main, Germany

Maria J. G. T. Vehreschild & Yascha Khodamoradi

299. Humanitas Gavazzeni-Castelli, Bergamo, Italy

Massimo Castoldi & Stefano Duga

300. School of Biological Sciences, Monash University, Clayton, Victoria, Australia

Mauro D' Amato & Tenghao Zheng

301. Munich Clinic Schwabing, Academic Teaching Hospital, Ludwig-Maximilians-University (LMU), Munich, Germany

Michael J. Seilmaier

302. Department of Anesthesiology, Hospital Universitario Ramón y Cajal,  
Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid,  
Spain

Nilda Martínez

303. Clinical Trials Centre Cologne, ZKS Köln, Cologne, Germany

Oliver A. Cornely

304. Institute of Human Genetics, University of Bonn School of Medicine,  
University Hospital Bonn, Bonn, Germany

Per Hoffmann & Stefanie Heilmann-Heimbach

305. Institute of Clinical Molecular Biology, Christian-Albrechts-University  
of Kiel, Kiel, Germany

Petra Bacher

306. UKSH Schleswig-Holstein, Kiel, Germany

Petra Bacher

307. Institute of Immunology, Christian-Albrechts-University of Kiel, Kiel,  
Germany

Petra Bacher

308. Institute of Medical Virology, University Hospital Frankfurt, Goethe  
University, Frankfurt am Main, Germany

Sandra Ciesek

309. German Centre for Infection Research (DZIF), External Partner Site  
Frankfurt, Frankfurt am Main, Germany

Sandra Ciesek

310. Department of Neurology, Bezirksklinikum Regensburg, University of Regensburg, Regensburg, Germany

Sibylle Wilfling

311. Charite Universitätsmedizin Berlin, Berlin Institute of Health, Berlin, Germany

Ulf Landmesser

312. German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany

Ulrike Protzer

313. Department of Infectious Diseases, University Hospital of North Norway, Tromsø, Norway

Vegard Skogen

314. Faculty of Health Sciences, UIT The Arctic University of Norway, Tromsø, Norway

Vegard Skogen

315. Catalan Institute of Oncology (ICO), Barcelona, Spain

Victor Moreno

316. Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain

Victor Moreno

317. Universitat de Barcelona (UB), Barcelona, Spain

Victor Moreno

318. deCODE genetics, Reykjavik, Iceland

Enrique J. Calderón, Francisco J. Medrano, Juan Delgado, Rubén Morilla, Vicente Friaza, Daniel F. Gudbjartsson, Kari Stefansson, Patrick Sulem, Gardar Sveinbjornsson, Pall Melsted, Gudmundur Norddahl, Kristjan Helgi Swerford Moore, Unnur Thorsteinsdottir & Hilma Holm

319. Mucosal Immunology Lab, Unidad de Excelencia Instituto de Biomedicina y Genética Molecular de Valladolid (IBGM), Universidad de Valladolid-CSIC, Valladolid, Spain

David Bernardo

320. Centro de Investigaciones Biomédicas en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid, Spain

Javier Ampuero & David Bernardo

321. Valladolid University Hospital, Valladolid, Spain

Silvia Rojo Rello

322. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Reedik Magi, Lili Milani, Andres Metspalu, Triin Laisk, Kristi Läll, Maarja Lepamets, Tõnu Esko, Ene Reimann, Helene Alavere, Kristjan Metsalu & Mairo Puusepp

323. SYNLAB Estonia, University of Tartu, Tartu, Estonia

Paul Naaber

324. University of Tartu, Tartu, Estonia

Edward Laane

325. Kuressaare Hospital, Kuressaare, Estonia

Edward Laane & Jaana Pesukova

326. Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

Pärt Peterson & Kai Kisand

327. West Tallinn Central Hospital, Tallinn, Estonia

Jekaterina Tabri, Raili Allos & Kati Hensen

328. University of Tartu, Tartu University Hospital, Tartu, Estonia

Joel Starkopf

329. Estonian Health Insurance Fund, Tallinn, Estonia

Inge Ringmets

330. Tartu University Hospital, Tartu, Estonia

Anu Tamm & Anne Kallaste

331. FinnGen, Helsinki, Finland

Aarno Palotie

332. Finnish Institute for Health and Welfare (THL), Helsinki, Finland

Kati Kristiansson, Sami Koskelainen & Markus Perola

333. University of Helsinki, Faculty of Medicine, Clinical and Molecular Metabolism Research Program, Helsinki, Finland

Markus Perola

334. Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland

Carlo Rivolta, Mathieu Quinodoz & Dhryata Kamdar

335. Department of Ophthalmology, University of Basel, Basel, Switzerland

Carlo Rivolta, Mathieu Quinodoz & Dhryata Kamdar

336. Infectious Diseases Service, Department of Medicine, University Hospital and University of Lausanne, Lausanne, Switzerland

Pierre-Yves Bochud, Stéphanie Bibert & Noémie Boillat

337. Centre for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland

Semira Gonseth Nussle

338. Division of Infectious Diseases and Hospital Epidemiology, Cantonal Hospital St Gallen, St Gallen, Switzerland

Werner Albrich

339. Division of Intensive Care, Geneva University Hospitals and the University of Geneva Faculty of Medicine, Geneva, Switzerland

Noémie Suh

340. Infectious Disease Service, Department of Internal Medicine, Geneva University Hospital, Geneva, Switzerland

Dionysios Neofytos

341. Clinique de Médecine et spécialités, Infectiologie, HFR-Fribourg, Fribourg, Switzerland

Véronique Erard

342. Infectious Diseases Division, University Hospital Centre of the Canton of Vaud, Hospital of Valais, Sion, Switzerland

Cathy Voide

343. Functional Host Genomics of Infectious Diseases, University Hospital and University of Lausanne, Lausanne, Switzerland

P. Y. Bochud, C. Rivolta, S. Bibert, M. Quinodoz, D. Kamdar, D. Neofytos, V. Erard, C. Voide, R. Friolet, P. Vollenweider, J. L. Pagani, M. Oddo, F. Meyer zu Bentrup, A. Conen, O. Clerc, O. Marchetti, A. Guillet, C. Guyat-Jacques, S. Foucras, M. Rime, J. Chassot, M. Jaquet, R. Merlet Viollet, Y. Lannepoudenx, L. Portopena & P. Y. Bochud

344. Registry COVID, University Hospital and University of Lausanne, Lausanne, Switzerland

P. Y. Bochud, P. Vollenweider, J. L. Pagani, F. Desgranges, P. Filippidis, B. Guéry, D. Haefliger, E. E. Kampouri, O. Manuel, A. Munting, M. Papadimitriou-Olivgeris, J. Regina, L. Rochat-Stettler, V. Suttels, E. Tadini, J. Tschopp, M. Van Singer, B. Viala & P. Y. Bochud

345. Pneumonia Prediction using Lung Ultrasound, University Hospital and University of Lausanne, Lausanne, Switzerland

N. Boillat-Blanco, T. Brahier, O. Hügli, J. Y. Meuwly & O. Pantet

346. Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

Tomas Pumarola, S. Gonseth Nussle, M. Bochud, V. D'Acremont & S. Estoppey Younes

347. COVID-19 Risk Prediction in Swiss ICUs-Trial, Division of Infectious Diseases and Hospital Epidemiology, Cantonal Hospital St Gallen, St Gallen, Switzerland

P. Y. Bochud, W. C. Albrich, N. Suh, A. Cerny, L. O'Mahony, P. Y. Bochud, M. Frischknecht, G.-R. Kleger, M. Filipovic, C. R. Kahlert, H. Wozniak, T. Rochat Negro, J. Pugin, K. Bouras, C. Knapp, T. Egger, A. Perret, P. Montillier, C. di Bartolomeo & B. Barda

348. GCAT-Genomes for Life, Germans Trias i Pujol Health Sciences Research Institute (IGTP), Badalona, Spain

Rafael de Cid, Anna Carreras, Iván Galván-Femenía, Natalia Blay, Xavier Farré, Lauro Sumoy & Beatriz Cortés

349. Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, Consortium for Biomedical Research in Epidemiology and Public Health, University of Barcelona, Barcelona, Spain

Victor Moreno

350. Barcelona Supercomputing Center, Centro Nacional de Supercomputación (BSC-CNS), Life & Medical Sciences, Barcelona, Spain

Josep Maria Mercader, Marta Guindo-Martinez & David Torrents

351. University of Siena, DIISM-SAILAB, Siena, Italy

Marco Gori & Nicola Picchiotti

352. Université Côte d'Azur, Inria, CNRS, I3S, Maasai, Nice, France

Marco Gori

353. Division of Infectious Diseases and Immunology, Department of Medical Sciences and Infectious Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

Mario Umberto Mondelli, Raffaele Bruno & Serena Ludovisi

354. Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy

Mario Umberto Mondelli, Raffaele Bruno & Serena Ludovisi

355. Department of Infectious and Tropical Diseases, University of Brescia and ASST Spedali Civili Hospital, Brescia, Italy

Francesco Castelli, Eugenia Quiros-Roldan & Melania Degli Antoni

356. Chirurgia Vascolare, Ospedale Maggiore di Crema, Crema, Italy

Massimo Vaghi

357. III Infectious Diseases Unit, ASST-FBF-Sacco, Milan, Italy

Stefano Rusconi & Agostino Riva

358. Department of Biomedical and Clinical Sciences Luigi Sacco,  
University of Milan, Milan, Italy

Stefano Rusconi, Matteo Siano, Arianna Gabrieli & Agostino Riva

359. Department of Specialized and Internal Medicine, Tropical and  
Infectious Diseases Unit, Azienda Ospedaliera Universitaria Senese,  
Siena, Italy

Francesca Montagnani, Mario Tumbarello, Massimiliano  
Fabbiani, Barbara Rossetti, Giacomo Zanelli & Ilaria Rancan

360. Unit of Respiratory Diseases and Lung Transplantation, Department of  
Internal and Specialist Medicine, University of Siena, Siena, Italy

Elena Bargagli, Laura Bergantini, Miriana D'Alessandro, Paolo  
Cameli & David Bennet

361. Department of Emergency and Urgency, Medicine, Surgery and  
Neurosciences, Unit of Intensive Care Medicine, Siena University  
Hospital, Siena, Italy

Federico Franchi, Federico Anedda, Simona Marcantonio & Sabino  
Scolletta

362. Department of Medical, Surgical and Neurosciences and Radiological  
Sciences, Unit of Diagnostic Imaging, University of Siena, Siena, Italy

Maria Antonietta Mazzei & Susanna Guerrini

363. Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Policlinico Le Scotte, Siena, Italy

Luca Cantarini, Edoardo Conticini & Bruno Frediani

364. Department of Specialized and Internal Medicine, Infectious Diseases Unit, San Donato Hospital Arezzo, Arezzo, Italy

Danilo Tacconi & Chiara Spertilli

365. Department of Emergency, Anesthesia Unit, San Donato Hospital, Arezzo, Italy

Marco Feri & Alice Donati

366. Department of Specialized and Internal Medicine, Pneumology Unit and UTIP, San Donato Hospital, Arezzo, Italy

Raffaele Scala & Luca Guidelli

367. Department of Emergency, Anesthesia Unit, Misericordia Hospital, Grosseto, Italy

Genni Spargi & Marta Corridi

368. Department of Specialized and Internal Medicine, Infectious Diseases Unit, Misericordia Hospital, Grosseto, Italy

Cesira Nencioni & Leonardo Croci

369. Department of Preventive Medicine, Azienda USL Toscana Sud Est, Arezzo, Italy

Maria Bandini, Maurizio Spagnesi, Paolo Piacentini, Elena Desanctis & Silvia Cappelli

370. Clinical Chemical Analysis Laboratory, Misericordia Hospital, Grosseto, Italy

Gian Piero Caldarelli

371. Territorial Scientific Technician Department, Azienda USL Toscana Sud Est, Arezzo, Italy

Anna Canaccini, Agnese Verzuri & Valentina Anemoli

372. Clinical Chemical Analysis Laboratory, San Donato Hospital, Arezzo, Italy

Agostino Ognibene, Alessandro Pancrazi & Maria Lorubbio

373. Department of Health Sciences, Clinic of Infectious Diseases, ASST Santi Paolo e Carlo, University of Milan, Milan, Italy

Antonella D' Arminio Monforte, Esther Merlini & Federica Gaia Miraglia

374. Department of Anesthesia and Intensive Care, University of Modena and Reggio Emilia, Modena, Italy

Massimo Girardis, Stefano Busani & Sophie Venturelli

375. HIV/AIDS Department, National Institute for Infectious Diseases, IRCCS, Lazzaro Spallanzani, Rome, Italy

Andrea Antinori, Arianna Emiliozzi & Alessandra Vergori

376. Infectious Diseases Clinic, Department of Medicine, Azienda Ospedaliera di Perugia, Perugia, Italy

Daniela Francisci & Elisabetta Schiaroli

377. Infectious Diseases Clinic, Santa Maria Hospital, University of Perugia, Perugia, Italy

Daniela Francisci, Elisabetta Schiaroli & Francesco Paciosi

378. Department of Infectious Diseases, Treviso Hospital, Treviso, Italy

Pier Giorgio Scotton & Francesca Andretta

379. Clinical Infectious Diseases, Mestre Hospital, Venezia, Italy

Sandro Panese

380. Infectious Diseases Clinic, ULSS1, Belluno, Italy

Renzo Scaggiante & Francesca Gatti

381. Medical Genetics and Laboratory of Medical Genetics Unit, A.O.R.N  
“Antonio Cardarelli”, Naples, Italy

Matteo Della Monica & Carmelo Piscopo

382. Department of Molecular Medicine and Medical Biotechnology,  
University of Naples Federico II, Naples, Italy

Mario Capasso, Roberta Russo, Immacolata Andolfo, Achille  
Iolascon & Giuseppe Merla

383. CEINGE Biotecnologie Avanzate, Naples, Italy

Mario Capasso, Roberta Russo, Immacolata Andolfo & Achille  
Iolascon

384. IRCCS SDN, Naples, Italy

Mario Capasso

385. Unit of Respiratory Physiopathology, AORN dei Colli, Monaldi  
Hospital, Naples, Italy

Giuseppe Fiorentino

386. Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della  
Sofferenza Hospital, San Giovanni Rotondo, Italy

Marco Castori & Massimo Carella

387. Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy

Filippo Aucella

388. Infectious Diseases Clinic, Policlinico San Martino Hospital, IRCCS for Cancer Research, Genova, Italy

Antonio Di Biagio & Matteo Bassetti

389. Microbiology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Catholic University of Medicine, Rome, Italy

Luca Masucci, Alessandra Guarnaccia & Maurizio Sanguinetti

390. Department of Laboratory Sciences and Infectious Diseases, Fondazione Policlinico Universitario AGemelli IRCCS, Rome, Italy

Luca Masucci, Alessandra Guarnaccia & Maurizio Sanguinetti

391. Department of Cardiovascular Diseases, University of Siena, Siena, Italy

Serafina Valente & Oreste De Vivo

392. Otolaryngology Unit, University of Siena, Siena, Italy

Marco Mandalà, Alessia Giorli & Lorenzo Salerni

393. Department of Internal Medicine, ASST Valtellina e Alto Lario, Sondrio, Italy

Patrizia Zucchi & Pierpaolo Parravicini

394. First Aid Department, Luigi Curto Hospital, Polla, Italy

Ferdinando Giannattasio, Tullio Trotta & Gabriella Coiro

395. U.O.C. Laboratorio di Genetica Umana, Genova, Italy

Domenico A. Coviello

396. Infectious Diseases Clinics, University of Modena and Reggio Emilia, Modena, Italy

Cristina Mussini

397. Department of Respiratory Diseases, Azienda Ospedaliera di Cremona, Cremona, Italy

Giancarlo Bosio & Enrico Martinelli

398. U.O.C. Medicina, ASST Nord Milano, Ospedale Bassini, Milan, Italy

Luisa Tavecchia, Mary Ann Belli & Sandro Mancarella

399. Department of Cardiovascular, Neural and Metabolic Sciences, Istituto Auxologico Italiano, IRCCS, San Luca Hospital, Milan, Italy

Lia Crotti & Gianfranco Parati

400. Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

Lia Crotti & Gianfranco Parati

401. Center for Cardiac Arrhythmias of Genetic Origin, Istituto Auxologico Italiano, IRCCS, Milan, Italy

Lia Crotti

402. Laboratory of Cardiovascular Genetics, Istituto Auxologico Italiano, IRCCS, Milan, Italy

Lia Crotti

403. Unit of Infectious Diseases, ASST Papa Giovanni XXIII Hospital, Bergamo, Italy

Marco Rizzi, Franco Maggiolo & Diego Ripamonti

404. Department of Cardiology, Institute of Montescano, Istituti Clinici Scientifici Maugeri, IRCCS, Pavia, Italy

Maria Teresa La Rovere

405. Department of Cardiac Rehabilitation, Institute of Tradate (VA), Istituti Clinici Scientifici Maugeri, IRCCS, Pavia, Italy

Simona Sarzi-Braga

406. Cardiac Rehabilitation Unit, Fondazione Salvatore Maugeri, IRCCS, Scientific Institute of Milan, Milan, Italy

Maurizio Bussotti

407. IRCCS CMondino Foundation, Pavia, Italy

Sabrina Ravaglia & Ravaglia Sabrina

408. Medical Genetics Unit, Meyer Children's University Hospital, Florence, Italy

Rosangela Artuso

409. Department of Medicine, Pneumology Unit, Misericordia Hospital, Grosseto, Italy

Antonio Perrella & Francesco Bianchi

410. Department of Preventive Medicine, Azienda USL Toscana Sud Est, Arezzo, Italy

Davide Romani

411. Department of Anesthesia and Intensive Care Unit, ASST Fatebenefratelli Sacco, Luigi Sacco Hospital, Polo Universitario, University of Milan, Milan, Italy

Paola Bergomi, Emanuele Catena & Riccardo Colombo

412. Health Management, Azienda USL Toscana Sudest, Arezzo, Italy

Simona Dei

413. Department of Mathematics, University of Pavia, Pavia, Italy

Nicola Picchiotti & Marco Tanfoni

414. Independent researcher, Milan, Italy

Maurizio Sanarico

415. Scuola Normale Superiore, Pisa, Italy

Francesco Raimondi

416. CNR-Consiglio Nazionale delle Ricerche, Istituto di Biologia e Biotecnologia Agraria (IBBA), Milano, Italy

Filippo Biscarini & Alessandra Stella

417. Veos Digital, Milan, Italy

Mattia Bergomi

418. Core Research Laboratory, ISPRO, Florence, Italy

Katia Capitani

419. Division of Infectious Diseases and Immunology, Fondazione IRCCS Polyclinico San Matteo, Pavia, Italy

Marco Vecchia & Stefania Mantovani

420. Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

Isabella Zanella

421. Clinical Chemistry Laboratory, Cytogenetics and Molecular Genetics Section, Diagnostic Department, ASST Spedali Civili di Brescia, Brescia, Italy

Isabella Zanella

422. Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy

Andrea Cossarizza

423. Department of Molecular Medicine, University of Padova, Padua, Italy

Saverio Giuseppe Parisi & Stefano Baratti

424. Laboratory of Regulatory and Functional Genomics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

Giuseppe Merla & Gabriella Maria Squeo

425. Clinical Trial Office, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy

Pamela Raggi, Carmen Marciano & Rita Perna

426. Department of Health Sciences, University of Genova, Genova, Italy

Matteo Bassetti

427. Oncologia Medica e Ufficio Flussi Sondrio, Sondrio, Italy

Elisabetta Menatti

428. Local Health Unit, Pharmaceutical Department of Grosseto, Toscana Sud Est Local Health Unit, Grosseto, Italy

Fabio Lena

429. Independent researcher, Milan, Italy

Chiara Gabbi

430. Direzione Scientifica, Istituti Clinici Scientifici Maugeri IRCCS,  
Pavia, Italy

Tiziana Bachetti

431. Fondazione per la ricerca Ospedale di Bergamo, Bergamo, Italy

Claudia Suardi

432. Allelica, New York, NY, USA

Giordano Bottà & Paolo Di Domenico

433. School of Basic and Medical Biosciences, Faculty of Life Sciences and  
Medicine, King's College London, London, UK

Richard C. Trembath

434. Medical and Population Genomics, Wellcome Sanger Institute,  
Hinxton, UK

Qin Qin Huang & Hilary C. Martin

435. Bradford Institute for Health Research, Bradford Teaching Hospitals  
National Health Service (NHS) Foundation Trust, Bradford, UK

Dan Mason & John Wright

436. Institute of Population Health Sciences, Queen Mary University of  
London, London, UK

Sarah Finer

437. Genes & Health, Blizzard Institute, Queen Mary University of London,  
London, UK

Shaheen Akhtar, Mohammad Anwar, Elena Arciero, Samina Ashraf, Gerome Breen, Raymond Chung, Charles J. Curtis, Maharun Chowdhury, Grainne Colligan, Panos Deloukas, Ceri Durham, Sarah Finer, Chris Griffiths, Qin Qin Huang, Matt Hurles, Karen A. Hunt, Shapna Hussain, Kamrul Islam, Ahsan Khan, Amara Khan, Cath Lavery, Sang Hyuck Lee, Robin Lerner, Daniel MacArthur, Bev MacLaughlin, Hilary Martin, Dan Mason, Shefa Miah, Bill Newman, Nishat Safa, Farah Tahmasebi, Richard C. Trembath, Bhavi Trivedi, David A. van Heel & John Wright

438. Institute of Population Health Sciences, Queen Mary University of London, London, UK

Christopher J. Griffiths

439. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Albert V. Smith, Andrew P. Boughton, Kevin W. Li, Jonathon LeFaive, Aubrey Annis, Sebastian Zöllner, Jiongming Wang & Andrew Beck

440. Heart Institute (InCor), University of Sao Paulo Med School, São Paulo, Brazil

Cinthia E. Jannes, Jose E. Krieger, Alexandre C. Pereira, Mariliza Velho, Emanuelle Marques, Isabella Ramos Lima, Mauricio Teruo Tada & Karina Valino

441. Genentech, San Francisco, CA, USA

Mark McCarthy, Carrie Rosenberger, Diana Chang, Christian Hammer, Julie Hunkapiller, Anubha Mahajan, Sarah Pendergrass, Lara Sucheston-Campbell & Brian Yaspan

442. DNA Link Inc., Seoul, Republic of Korea

Jong Eun Lee, Hyun Soo Lee, Eunsoon Shin & Hye Yoon Jang

443. Seoul National University Hospital Gangnam Center, Seoul, Republic of Korea

Sunmie Kim

444. Division of Infectious Diseases, Department of Internal Medicine, Chungnam National University School of Medicine, Daejeon, Republic of Korea

Sungmin Kym, Yeon-Sook Kim & Hyeongseok Jeong

445. East Kent Hospitals NHS Foundation Trust, Canterbury, UK

Ana Alegria

446. Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Ki Tae Kwon & Shin-Woo Kim

447. Division of Infectious Diseases, Department of Internal Medicine, Incheon Medical Center, Incheon, Republic of Korea

Jin Yong Kim & Young Rock Jang

448. Department of Infectious Diseases, Keimyung University Dongsan Hospital, Keimyung University School of Medicine, Daegu, Republic of Korea

Hyun ah Kim & Ji Yeon Lee

449. Department of Internal Medicine, Pusan National University School of Medicine and Medical Research Institute, Pusan National University Hospital, Busan, Republic of Korea

Jeong Eun Lee & Shinwon Lee

450. Division of Infectious Diseases, Department of Internal Medicine, Myongji Hospital, Goyang, Republic of Korea

Kang-Won Choe & Yu Min Kang

451. Institute for Health Promotion, Graduate School of Public Health,  
Yonsei University, Seoul, Republic of Korea

Sun Ha Jee & Keum Ji Jung

452. Division of Cardiovascular Medicine, Stanford University, Stanford,  
CA, USA

Victoria Parikh, Matthew Wheeler, Karen Dalton, Jeff Christle, John  
Gorzynski, Hannah de Jong, Shirley Sutton, Nathan Youlton, Ruchi  
Joshi, David Jimenez-Morales, Christopher Hughes, David  
Amar, Steve Hershman, Anna Kirillova, Kinya Seo, Yong  
Huang, Archana Raja & Jimmy Zhen

453. Department of Medicine, Stanford University, Stanford, CA, USA

Euan Ashley

454. Department of Genetics, Stanford University, Stanford, CA, USA

Euan Ashley & Carlos Bustamante

455. Department of Biomedical Data Science, Stanford University,  
Stanford, CA, USA

Manuel Rivas, Carlos Bustamante, Yosuke Tanigawa & Alex Ioannidis

456. Department of Pathology, Stanford University, Stanford, CA, USA

Benjamin Pinsky, Massa Shoura, Nathan Hammond, Nathaniel  
Watson, ChunHong Huang, Malaya Sahoo & Hannah Wang

457. Illumina, San Diego, CA, USA

Phillip Febbo, Kyle Farh, Gary P. Schroth & Francis deSouza

458. Computational Biology, Drug Discovery Sciences, Takeda Pharmaceuticals, Boston, MA, USA

Christopher Deboever & Sndor Szalma

459. Department of Computational Biology, Swiss Institute of Bioinformatics (SIB), University of Lausanne, Lausanne, Switzerland

Simone Rubinacci & Olivier Delaneau

460. Royal Victoria Hospital, Belfast, UK

Peter J. McGuigan, Christopher Wasson, Stephanie Finn, Jackie Green, Erin Collins & Bernadette King

461. Chelsea & Westminster NHS Foundation Trust, London, UK

Luke Stephen Prockter Moore, Marcela Paola Vizcaychipi, Laura Gomes de Almeida Martins & Jaime Carungcong

462. Northampton General Hospital NHS Trust, Northampton, UK

Kathryn Hall & Isheunesu Mapfunde

463. Wrexham Maelor Hospital, Wrexham, UK

Andy Campbell, Sara Smuts, Joseph Duffield, Oliver Smith, Lewis Mallon & Watkins Claire

464. University College Dublin, St Vincent's University Hospital, Dublin, Ireland

Ailstair Nichol, Kathy Brickell, Michelle Smyth & Lorna Murphy

465. University Hospitals Coventry & Warwickshire NHS Trust, Coventry, UK

Geraldine Ward & Pamela Bremmer

466. Watford General Hospital, Watford, UK

Valerie Joan Page & Siobhain Carmody

467. NIHR Health Protection Research Unit, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK

Malcolm G. Semple, Tom Solomon, Lance C. W. Turtle & Hayley Hardwick

468. Queen Alexandra Hospital (Hampshire), Portsmouth Hospital Trust, Portsmouth, UK

Kayode Adeniji

469. Princess Royal Hospital, Brighton & Sussex Universities Hospitals NHS Trust, Brighton, UK

Daniel Agranoff & Chi Eziefula

470. Bassetlaw Hospital, Doncaster and Bassetlaw, Worksop, UK

Ken Agwu

471. Darent Valley Hospital, Dartford & Gravesham NHS Trust, Dartford, UK

Dhiraj Ail

472. High Containment Laboratories, University of Birmingham, Birmingham, UK

Erin L. Aldera

473. Queen Elizabeth the Queen Mother Hospital, Margate, UK

Ana Alegria, Louise Allen, Eva Beranova, Nikki Crisp, Joanne Deery, Tracy Hazelton, Alicia Knight, Carly Price, Sorrell

Tilbey, Salah Turki & Sharon Turney

474. John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Brian Angus

475. Royal Albert Edward Infirmary (Wigan), Wrightington, Wigan and Leigh, Wigan, UK

Abdul Ashish

476. Manchester Royal Infirmary, Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Dougal Atkinson

477. Furness General Hospital, Morecambe Bay NHS Foundation Trust, Barrow-in-Furness, UK

Shahedal Bari

478. Castle Hill Hospital, Hull University Teaching Hospital Trust, Hull, UK

Gavin Barlow

479. Hillingdon Hospital, Hillingdon Hospital, London, UK

Stella Barnass

480. St Thomas Hospital, Guys and St Thomas Foundation Trust, London, UK

Nicholas Barrett, Sam Douthwaite, Marlies Ostermann & Manu Shankar-Hari

481. University Hospitals Coventry and Warwickshire, Coventry, UK

Christopher Bassford

482. St Michaels Hospital (Bristol), University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

Sneha Basude

483. Stepping Hill Hospital, Stockport NHS Foundation Trust, Manchester, UK

David Baxter

484. Royal Liverpool Hospital, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK

Michael Beadsworth & Ingeborg Welters

485. Bristol Royal Hospital (Children's), University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

Jolanta Bernatoniene

486. Scarborough Hospital, York Teaching Hospitals NHS Foundation Trust, York, UK

John Berridge

487. Liverpool Heart & Chest Hospital, Liverpool Heart & Chest NHS Foundation Trust, Liverpool, UK

Nicola Best

488. James Paget University Hospital, James Paget University Hospitals NHS Foundation Trust, Great Yarmouth, UK

Pieter Bothma & Darell Tupper-Carey

489. The James Cook University Hospital, South Tees NHS Foundation Trust, Middlesbrough, UK

David Chadwick

490. Aberdeen Royal Infirmary, Grampian, Aberdeen, UK

Robin Brittain-Long

491. University of Edinburgh, Edinburgh, UK

Naomi Bulteel, Lee Murphy, Nicola Wrobel, Sarah McCafferty, Kirstie Morrice & Alan MacLean

492. Royal Devon and Exeter Hospital, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Tom Burden

493. Worcestershire Royal Hospital, Worcestershire Acute Hospitals NHS Trust, Worcester, UK

Andrew Burtenshaw

494. Conquest Hospital, Hastings, East Sussex Healthcare NHS Trust, Seaford, UK

Vikki Caruth

495. Dorset County Hospital, Dorset County Hospital NHS Foundation Trust, Dorchester, UK

Duncan Chamberlain

496. Royal Bournemouth General Hospital, University Hospitals Dorset NHS Foundation Trust, Bournemouth, UK

Nigel Chee

497. Harrogate Hospital, Harrogate and District NHS Foundation Trust, Harrogate, UK

Jenny Child

498. Burnley General Teaching Hospital, East Lancashire Hospitals NHS Hospitals, Burnley, UK

Srikanth Chukkambotla

499. Torbay Hospital, Torbay & South Devon NHS Foundation Trust, Torquay, UK

Tom Clark

500. Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Paul Collini, Cariad Evans, Gary H. Mills, Norfaizan Ahmad, Joann Barker, Kris Bauchmuller, Sarah Bird, Kay Cawthon, Kate Harrington, Yvonne Jackson, Faith Kibutu, Becky Lenagh, Shamiso Masuko, Gary H. Mills, Ajay Raithatha, Matthew Wiles, Jayne Willson, Helen Newell, Alison Lye, Lorenza Nwafor, Claire Jarman, Sarah Rowland-Jones, David Foote, Joby Cole, Roger Thompson, James Watson, Lisa Hesseldon, Irene Macharia, Luke Chetam, Jacqui Smith, Amber Ford, Samantha Anderson, Kathryn Birchall, Kay Housley, Sara Walker, Leanne Milner, Helena Hanratty, Helen Trower, Patrick Phillips, Simon Oxspring & Ben Donne

501. St Georges Hospital (Tooting), St Georges University Hospitals NHS Foundation Trust, London, UK

Catherine Cosgrove

502. Blackpool Victoria Hospital, Blackpool Teaching Hospitals NHS Foundation Trust, Blackpool, UK

Jason Cupitt

503. The Royal London Hospital, Barts Health NHS Trust, London, UK

Maria-Teresa Cutino-Moguel

504. Salford Royal NHS Foundation Trust, Salford Royal NHS Foundation Trust, Manchester, UK

Paul Dark

505. University Hospital of North Durham, County Durham and Darlington Foundation Trust, Durham, UK

Chris Dawson

506. Norfolk and Norwich University Hospital, Norfolk and Norwich University Hospital NHS Foundation Trust, Norwich, UK

Samir Dervisevic

507. Fairfield General Hospital, Pennine Acute Hospitals NHS Trust, Manchester, UK

Andrew Drummond & Andrew Ustianowski

508. Hereford County Hospital, Wye Valley NHS Trust, Hereford, UK

Ingrid DuRand

509. Southampton General Hospital, University Hospital Southampton NHS Foundation Trust, Southampton, UK

Ahilanadan Dushianthan

510. Northampton General Hospital, Northampton General Hospital NHS Trust, Northampton, UK

Tristan Dyer

511. University Hospital of Wales, Cardiff and Vale University Health Board, Cardiff, UK

Christopher Fegan

512. University of Bristol, Bristol, UK

Adam Finn

513. Leighton Hospital, Mid Cheshire Hospitals NHS Foundation Trust, Crewe, UK

Duncan Fullerton & Elijah Matovu

514. Diana Princess of Wales Hospital (Grimsby), North Lincolnshire & Goole, Grimsby, UK

Sanjeev Garg

515. Manor Hospital, Walsall Healthcare NHS Trust, Walsall, UK

Atul Garg

516. Addenbrookes Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK

Effrossyni Gkrania-Klotsas & Elinoor Moore

517. West Suffolk Hospital, West Suffolk Hospital NHS Foundation Trust, Bury St Edmunds, UK

Jo Godden

518. Basingstoke and North Hampshire Hospital, Hampshire Hospitals NHS Foundation Trust, Basingstoke, UK

Arthur Goldsmith

519. North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK

Clive Graham

520. Warwick Hospital, South Warwickshire NHS Foundation Trust,  
Warwick, UK

Elaine Hardy

521. Birmingham Women's and Children's Hospital, Birmingham Women's  
and Children's Hospital NHS Foundation Trust, Birmingham, UK

Stuart Hartshorn

522. Nottingham City Hospital, Nottingham University Hospitals NHS  
Trust, Nottingham, UK

Daniel Harvey

523. Glangwili Hospital Child Health Section, Hywel Dda University  
Health Board, Carmarthen, UK

Peter Havalda & Andrew Workman

524. Alder Hey Children's Hospital, Alder Hey Children's NHS Foundation  
Trust, Liverpool, UK

Daniel B. Hawcutt

525. Bronglais General Hospital, Hywel Dda University Health Board,  
Aberystwyth, UK

Maria Hobrok

526. Worthing Hospital, Western Sussex Hospitals NHS Foundation Trust,  
Worthing, UK

Luke Hodgson

527. Rotheram District General Hospital, The Rotherham NHS Foundation  
Trust, Rotherham, UK

Anil Hormis

528. Royal Free Hospital, Royal Free London NHS Foundation Trust,  
London, UK

Michael Jacobs

529. Homerton Hospital, Homerton University Hospital NHS Foundation  
Trust, London, UK

Susan Jain

530. Airedale Hospital, Airedale NHS Foundation Trust, Keighley, UK

Paul Jennings

531. Basildon Hospital, Basildon and Thurrock University Hospitals NHS  
Foundation Trust, Basildon, UK

Agilan Kaliappan

532. The Christie NHS Foundation Trust, Manchester, UK

Vidya Kasipandian

533. Queen Elizabeth Hospital (Greenwich), Lewisham and Greenwich  
NHS Trust, London, UK

Stephen Kegg

534. The Whittington Hospital, Whittington Health NHS Trust, London,  
UK

Michael Kelsey

535. Sheffield Children's Hospital, Sheffield Children's NHS Foundation  
Trust, Sheffield, UK

Caroline Kerrison

536. Royal United Hospital, Bath, Royal United Hospitals Bath NHS Foundation Trust, Bath, UK

Ian Kerslake

537. Western General Hospital, Edinburgh, UK

Oliver Koch, Rosie Baruah, Sheila Morris, Susie Ferguson & Amy Shepherd

538. Mid and South Essex NHS Foundation Trust, Basildon, UK

Gouri Koduri

539. Hinchingbrooke Hospital, North West Anglia NHS Foundation Trust, Peterborough, UK

George Koshy, Tamas Leiner & Katherine Mortimore

540. Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

Shondipon Laha

541. University Hospital (Coventry), University Hospitals Coventry and Warwickshire, Coventry, UK

Steven Laird

542. The Walton Centre, The Walton Centre, Liverpool, UK

Susan Larkin

543. Hull Royal Infirmary, Hull University Teaching Hospital Trust, Hull, UK

Patrick Lillie

544. Darlington Memorial Hospital, County Durham and Darlington Foundation Trust, Darlington, UK

James Limb

545. Queen Elizabeth Hospital (Gateshead), Gateshead NHS Foundation Trust, Newcastle, UK

Vanessa Linnett

546. Warrington Hospital, Warrington & Halton Hospitals NHS Foundation Trust, Warrington, UK

Jeff Little

547. University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

Mark Lyttle

548. St Mary's Hospital (Isle of Wight), Isle of Wight NHS Trust, Isle of Wight, UK

Emily MacNaughton

549. The Maidstone Hospital, Maidstone & Tunbridge Wells NHS Trust, Maidstone, UK

Ravish Mankregod

550. Huddersfield Royal, Calderdale and Huddersfield NHS Foundation Trust, Huddersfield, UK

Huw Masson

551. Royal Surrey County Hospital, Guildford, UK

Katherine McCullough

552. Countess of Chester Hospital, Countess of Chester Hospital NHS Foundation Trust, Chester, UK

Ruth McEwen & Lawrence Wilson

553. Frimley Park Hospital, Frimley Health Foundation Trust, Frimley, UK

Manjula Meda

554. Leeds General Infirmary, Leeds Teaching Hospitals, Leeds, UK

Jane Minton & Karl Ward

555. North Middlesex Hospital, North Middlesex University Hospital NHS Trust, London, UK

Mariyam Mirfenderesky

556. Arrowe Park Hospital, Wirral University Teaching Hospital NHS Foundation Trust, Wirral, UK

Kavya Mohandas

557. Great Ormond Street Hospital, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Quen Mok

558. Royal Shrewsbury Hospital, Shrewsbury and Telford Hospital NHS Trust, Shrewsbury, UK

James Moon, Nigel Capps & Sanal Jose

559. East Surrey Hospital (Redhill), Surrey & Sussex Healthcare, Redhill, UK

Patrick Morgan

560. Burton Hospital, University Hospitals of Derby & Burton NHS Foundation Trust, Burton-on-Trent, UK

Craig Morris

561. Kent and Canterbury Hospital, East Kent Hospitals NHS Foundation Trust, Canterbury, UK

Samuel Moses

562. Weston Area General Trust, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

Mbiye Mpenge

563. Luton and Dunstable University Hospital, Luton, UK

Rohinton Mulla

564. Glasgow Royal Infirmary, Greater Glasgow and Clyde, Glasgow, UK

Michael Murphy

565. Derbyshire Healthcare, Derbyshire Healthcare NHS Foundation Trust, Derby, UK

Megan Nagel

566. Macclesfield General Hospital, East Cheshire NHS Foundation Trust, Macclesfield, UK

Thapas Nagarajan

567. Chelsea and Westminster Hospital, Chelsea and Westminster NHS Trust, London, UK

Mark Nelson

568. Institute of Microbiology and Infection, University of Birmingham,  
Birmingham, UK

Matthew K. O’Shea & Christoper A. Green

569. Prince Philip Hospital, Hwyl Dda University Health Board, Llanelli,  
UK

Igor Otahal

570. George Eliot Hospital - Acute Services, George Eliot Hospital,  
Nuneaton, UK

Mark Pais

571. Kettering General Hospital, Kettering General Hospital NHS  
Foundation Trust, Kettering, UK

Selva Panchatsharam

572. Birmingham Heartlands Hospital, Birmingham, UK

Danai Papakonstantinou, Hollie Bancroft, Mary Bellamy, Margaret  
Carmody, Jacqueline Daglish, Faye Moore, Joanne Rhodes, Mirriam  
Sangombe, Salma Kadiri & James Scriven

573. Russells Hall Hospital, The Dudley Group NHS Foundation Trust,  
Dudley, UK

Hassan Paraiso

574. Harefield Hospital, Royal Brompton & Harefield Trust, London, UK

Brij Patel

575. Lister Hospital, East and North Hertfordshire NHS Trust, Stevenage,  
UK

Natalie Pattison

576. Musgrove Park Hospital (Taunton & Somerset), Somerset NHS Foundation Trust, Taunton, UK

1. Justin Pepperell
2. Queen's Hospital, Havering (Romford), Barking, Havering and Redbridge University Hospitals NHS Trust, London, UK  
Mandeep Phull
3. Southport & Formby District General Hospital, Southport and Ormskirk Hospital NHS Trust, Southport, UK  
Stefania Pintus
4. New Cross Hospital, The Royal Wolverhampton NHS Trust, Wolverhampton, UK  
Jagtur Singh Pooni
5. King's College Hospital, London, UK  
Frank Post, Anna Cavazza, Maeve Cockrell, Eleanor Corcoran, Maria Depante, Clare Finney, Ellen Jerome, Mark McPhail, Monalisa Nayak, Harriet Noble, Kevin O'Reilly, Evita Pappa, Rohit Saha, Sian Saha, John Smith & Abigail Knighton
6. The Royal Victoria Infirmary, Newcastle Hospitals NHS Trust, Newcastle, UK  
David Price
7. The Great Western Hospital, Great Western Hospitals NHS Foundation Trust, Swindon, UK  
Rachel Prout
8. Ninewells Hospital, Tayside, Dundee, UK  
Nikolas Rae

9. Poole Hospital NHS Trust, Poole, UK

Henrik Reschreiter

10. Burton Hospital, University Hospitals of Derby & Burton NHS Foundation Trust, Derby, UK

Tim Reynolds

11. William Harvey Hospital, Ashford, East Kent Hospitals NHS Foundation Trust, Willesborough, UK

Neil Richardson

12. King's Mill Hospital, Sherwood Forest Hospitals NHS Foundation Trust, Sutton-in-Ashfield, UK

Mark Roberts

13. Liverpool Women's NHS Foundation Trust, Liverpool, UK

Devender Roberts

14. Dewsbury Hospital, Mid Yorkshire Hospitals NHS Trust, Dewsbury, UK

Alistair Rose

15. Northern Devon District Hospital, Northern Devon Healthcare NHS Trust, Barnstaple, UK

Guy Rousseau

16. Tameside General Hospital, Tameside and Glossop Integrated Care NHS Foundation Trust, Manchester, UK

Brendan Ryan

17. Sandwell General Hospital, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

Taranprit Saluja

18. Broomfield Hospital, Mid and South Essex University Hospitals Group, Broomfield, UK

Aarti Shah

19. Wycombe Hospital, Buckingham Healthcare NHS Trust, Wycombe, UK

Prad Shanmuga

20. University Hospital of North Tees, North Tees and Hartlepool NHS Trust, Stockton-on-Tees, UK

Anil Sharma

21. Royal Manchester Children's Hospital, Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Anna Shawcross

22. Bedford Hospital, Bedford, UK

Jeremy Sizer, Victoria Bastion, Daphene Clarke, Beena David, Harriet Kent, Rachel Lorusso, Gamu Lubimbi, Sophie Murdoch, Melchizedek Penacerrada, Alastair Thomas, Jennifer Valentine, Ana Vochin, Retno Wulandari & Brice Djeugam

23. Colchester General Hospital, East Suffolk and North Essex Foundation Trust, Colchester, UK

Richard Smith

24. Queen Elizabeth Hospital (Birmingham) and Heartlands, University Hospital Birmingham NHS Foundation Trust, Birmingham, UK

Catherine Snelson & Tony Whitehouse

25. Chesterfield Royal Hospital, Chesterfield Royal Hospital NHS Foundation Trust, Chesterfield, UK

Nick Spittle

26. Princess Alexandra Hospital, The Princess Alexandra Hospital NHS Trust, Harlow, UK

Nikki Staines & Shico Visuvanathan

27. Watford General Hospital, West Hertfordshire Hospitals NHS Trust, Watford, UK

Tom Stambach

28. Milton Keynes Hospital, Milton Keynes University Hospital NHS Foundation Trust, Milton Keynes, UK

Richard Stewart

29. Royal Bolton General Hospital, Bolton Foundation Trust, Bolton, UK

Pradeep Subudhi

30. Royal Gwent (Newport), Aneurin Bevan University Health Board, Newport, UK

Tamas Szakmany

31. The Royal Marsden Hospital (London), The Royal Marsden NHS Foundation Trust, London, UK

Kate Tatham

32. Queen Victoria Hospital (East Grinstead), Queen Victoria Hospital NHS Foundation Trust, East Grinstead, UK

Jo Thomas

33. County Hospital (Stafford), University Hospitals of North Midlands NHS Trust, Stafford, UK

Chris Thompson

34. Whiston Hospital, St Helen's & Knowlsey Hospitals NHS Trust, Prescot, UK

Ascanio Tridente

35. Croydon University Hospital, London, UK

Mary Twagira

36. Gloucester Royal, Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK

Nick Vallotton

37. Medway Maritime Hospital, Medway Maritime NHS Trust, Gillingham, UK

Lisa Vincent-Smith

38. Royal Papworth Hospital Everard, Royal Papworth Hospital NHS Foundation Trust, Cambridge, UK

Alan Vuylsteke

39. Derriford (Plymouth), University Hospital Plymouth NHS Trust, Plymouth, UK

Sam Waddy

40. St Helier Hospital, Epsom and St Helier University Hospital NHS Trust, London, UK

Rachel Wake

41. Royal Berkshire Hospital, Royal Berkshire Foundation Trust, London, UK

Andrew Walden

42. Bradford Royal Infirmary, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

Paul Whittaker

43. Northwick Park, London North West University Hospital Trust, London, UK

Ashley Whittington

44. Ealing Hospital, London North West University Hospital Trust, London, UK

Padmasayee Papineni

45. Royal Cornwall Hospital (Tresliske), Royal Cornwall NHS Trust, Truro, UK

Meme Wijesinghe

46. Ashford Hospital, Ashford & St Peter's Hospital, Stanwell, UK

Stephen Winchester

47. Leicester Royal Infirmary (includes Glenfield Site), University Hospitals of Leicester, Leicester, UK

Martin Wiselka

48. Grantham and District Hospital, United Lincolnshire Hospitals NHS Trust, Grantham, UK

Adam Wolverson

49. University Hospital Aintree, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK

Daniel G. Wooton

50. North Tyneside General Hospital, Northumbria Healthcare NHS Trust, North Shields, UK

Bryan Yates

51. Queen Elizabeth Hospital (King's Lynn), Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, King's Lynn, UK

Peter Young

52. The Crick Institute, London, UK

Rupert Beale

53. William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

Mark J. Caulfield & Charles Hinds

54. Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK

Elvina Gountouna, Caroline Hayward & David J. Porteous

55. Intensive Care National Audit & Research Centre, London, UK

David Harrison & Kathy Rowan

56. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

Gavin Band, Paul Klenerman & Julian Knight

57. Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK

Clark D. Russell, Debby Bogaert & Clark D. Russell

58. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Richard H. Scott

59. Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

Xia Shen, Zhijian Yang, Ranran Zhai & Chenqing Zheng

60. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, Edinburgh, UK

Xia Shen, Albert Tenesa & James F. Wilson

61. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Cynthia M. Bulik, Mikael Landén, Bengt Fundín & Xia Shen

62. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

Yang Wu

63. School of Life Sciences, Westlake University, Hangzhou, China

Jian Yang

64. Westlake Laboratory of Life Sciences and Biomedicine, Westlake University, Hangzhou, China

Jian Yang

65. Centre for Medical Informatics, The Usher Institute, University of Edinburgh, Edinburgh, UK

Lisa Norman, Riinu Pius, Thomas M. Drake, Cameron J. Fairfield, Stephen R. Knight, Kenneth A. Mclean, Derek Murphy, Catherine A. Shaw, Annemarie B. Docherty & Ewen M. Harrison

66. Liverpool Clinical Trials Centre, University of Liverpool, Liverpool, UK

Jo Dalton, Michelle Girvan, Egle Saviciute, Stephanie Roberts, Janet Harrison, Laura Marsh, Marie Connor, Sophie Halpin, Clare Jackson, Carrol Gamble & Chloe Donohue

67. Centre for Health Informatics, Division of Informatics, Imaging and Data Science, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

Gary Leeming

68. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Murray Wham

69. School of Informatics, University of Edinburgh, Edinburgh, UK

James Scott-Brown, Beatrice Alex & Benjamin Bach

70. Royal Hospital for Children, Glasgow, UK

Colin Begg

71. MRC-University of Glasgow Centre for Virus Research, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary

and Life Sciences, University of Glasgow, Glasgow, UK

Antonia Ying Wai Ho

72. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Peter W. Horby

73. Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China

Lowell Ling

74. Department of Critical Care Medicine, Queen's University and Kingston Health Sciences Centre, Kingston, Ontario, Canada

David Maslove

75. Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK

Danny McAuley

76. Department of Intensive Care Medicine, Royal Victoria Hospital, Belfast, UK

Danny McAuley

77. UCL Centre for Human Health and Performance, London, UK

Hugh Montgomery

78. Clinical Research Centre at St Vincent's University Hospital, University College Dublin, Dublin, Ireland

Alistair Nichol

79. National Heart and Lung Institute, Imperial College London, London, UK

Peter J. M. Openshaw & Ryan S. Thwaites

80. Imperial College Healthcare NHS Trust London, London, UK

Peter J. M. Openshaw

81. NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences University of Liverpool, Liverpool, UK

Malcolm G. Semple, Lance C. W. Turtle & Shona C. Moore

82. Respiratory Medicine, Alder Hey Children's Hospital, Institute in The Park, University of Liverpool, Alder Hey Children's Hospital, Liverpool, UK

Malcolm G. Semple

83. Department of Intensive Care Medicine, Guy's and St Thomas NHS Foundation Trust, London, UK

Manu Shankar-Hari

84. Department of Medicine, University of Cambridge, Cambridge, UK

Charlotte Summers

85. Airedale General Hospital, Keighley, UK

Lissa Armstrong, Hayley Bates, Emma Dooks, Fiona Farquhar, Brigid Hairsine, C. McParland & Sophie Packham

86. Barts Health NHS Trust, London, UK

Zoe Alldis, Raine Astin-Chamberlain, Fatima Bibi, Jack Biddle, Sarah Blow, Matthew Bolton, Catherine Borra, Ruth Bowles, Maudrian

Burton, Yasmin Choudhury, David Collier, Amber Cox, Amy Easthope, Patrizia Ebano, Stavros Fotiadis, Jana Gurasashvili, Rosslyn Halls, Pippa Hartridge, Delordson Kallon, Jamila Kassam, Ivone Lancoma-Malcolm, Maninderpal Matharu, Peter May, Oliver Mitchelmore, Tabitha Newman, Mital Patel, Jane Pheby, Irene Pinzuti, Zoe Prime, Oleksandra Prysyazhna, Julian Shiel, Melanie Taylor, Carey Tierney, Suzanne Wood, Anne Zak & Olivier Zongo

87. Basildon Hospital, Basildon, UK

Miranda Forsey, Agilan Kaliappan, Anne Nicholson, Joanne Riches & Mark Virtue

88. BHRUT (Barking Havering) - Queens Hospital and King George Hospital, Romford, UK

Lina Grauslyte, Musarat Hussain, Mandeep Phull, Tatiana Pogreban, Lace Rosaroso, Erika Salciute, George Franke, Joanna Wong & Aparna George

89. Bradford Royal Infirmary, Bradford, UK

Louise Akeroyd, Shereen Bano, Matt Bromley, Lucy Gurr, Tom Lawton, James Morgan, Kirsten Sellick, Deborah Warren, Brian Wilkinson, Janet McGowan, Camilla Ledgard, Amelia Stacey, Kate Pye, Ruth Bellwood & Michael Bentley

90. Bronglais General Hospital, Aberystwyth, UK

Maria Hobrok, Ronda Loosley, Heather McGuinness, Helen Tench & Rebecca Wolf-Roberts

91. Broomfield Hospital, Chelmsford, UK

Sian Gibson, Amanda Lyle, Fiona McNeela, Jayachandran Radhakrishnan & Alistair Hughes

92. Calderdale Royal Hospital, Halifax, UK

Asifa Ali, Megan Brady, Sam Dale, Annalisa Dance, Lisa Gledhill, Jill Greig, Kathryn Hanson, Kelly Holdroyd, Marie Home, Diane Kelly, Ross Kitson, Lear Matapure, Deborah Melia, Samantha Mellor, Tonicha Nortcliffe, Jez Pinnell, Matthew Robinson, Lisa Shaw, Ryan Shaw, Lesley Thomis, Alison Wilson, Tracy Wood, Lee-Ann Bayo, Ekta Merwaha, Tahira Ishaq & Sarah Hanley

93. Charing Cross Hospital, St Mary's Hospital and Hammersmith Hospital, London, UK

David Antcliffe, Dorota Banach, Stephen Brett, Phoebe Coghlan, Ziortza Fernandez, Anthony Gordon, Roceld Rojo, Sonia Sousa Arias & Maie Templeton

94. Barnet Hospital, London, UK

Rajeev Jha, Vinodh Krishnamurthy, Lai Lim & Li Lim

95. Birmingham Children's Hospital, Birmingham, UK

Rehana Bi, Barney Scholefield & Lydia Ashton

96. St John's Hospital Livingston, Livingston, UK

Alison Williams, Claire Cheyne, Anne Saunderson, Sam Moultrie & M. Odam

97. Aberdeen Royal Infirmary, Aberdeen, UK

Angela Allan, Felicity Anderson, Callum Kaye, Jade Liew, Jasmine Medhora, Teresa Scott, Erin Trumper & Adriana Botello

98. Addenbrooke's Hospital, Cambridge, UK

Petra Polgarova, Katerina Stroud, Eoghan Meaney, Megan Jones, Anthony Ng, Shruti Agrawal, Nazima Pathan, Deborah White, Esther Daubney & Kay Elston

99. Aintree University Hospital, Liverpool, UK

Robert Parker, Amie Reddy, Ian Turner-Bone, Laura Wilding & Peter Harding

100. Arrowe Park Hospital, Wirral, UK

Reni Jacob, Cathy Jones & Craig Denmade

101. Ashford and St Peter's Hospital, Lyne, UK

Maria Croft & Ian White

102. Basingstoke and North Hampshire Hospital, Basingstoke, UK

Denise Griffin, Nycola Muchenje, McDonald Mupudzi, Richard Partridge, Jo-Anna Conyngham, Rachel Thomas, Mary Wright & Maria Alvarez Corral

103. Borders General Hospital, Melrose, UK

Joy Dawson, Sweyn Garrioch, Melanie Tolson & Jonathan Aldridge

104. Chesterfield Royal Hospital Foundation Trust, Chesterfield, UK

Sarah Beavis, Katie Dale, Rachel Gascoyne, Joanne Hawes, Kelly Pritchard, Lesley Stevenson & Amanda Whileman

105. Eastbourne District General Hospital, East Sussex, UK and Conquest Hospital, Eastbourne, UK

Anne Cowley & Judith Highgate

106. Barnsley Hospital, Barnsley, UK

Rikki Crawley, Abigail Crew, Mishell Cunningham, Allison Daniels, Laura Harrison, Susan Hope, Ken Inwergbu, Sian Jones, Nicola Lancaster, Jamie Matthews, Alice Nicholson & Gemma Wray

107. Blackpool Victoria Hospital, Blackpool, UK

Leonie Benham, Zena Bradshaw, Joanna Brown, Melanie Caswell, Jason Cupitt, Sarah Melling, Stephen Preston, Nicola Slawson, Emma Stoddard & Scott Warden

108. East Surrey Hospital, Redhill, UK

Edward Combes, Teishel Joefield, Sonja Monnery, Valerie Beech & Sallyanne Trotman

109. Good Hope Hospital, Birmingham, UK

Bridget Hopkins, James Scriven, Laura Thrasyvoulou & Heather Willis

110. Hereford County Hospital, Hereford, UK

Susan Anderson, Janine Birch, Emma Collins, Kate Hammerton, Ryan O'Leary & Eleanor Andrews

111. Hull Royal Infirmary, Hull, UK

Caroline Abernathy, Louise Foster, Andrew Gratrix, Vicky Martinson, Priyai Parkinson, Elizabeth Stones & Llucia Carbral-Ortega

112. Kent & Canterbury Hospital, Canterbury, UK

Ritoo Kapoor, David Loader & Karen Castle

113. Manchester Royal Infirmary, Manchester, UK

Craig Brandwood, Lara Smith, Richard Clark, Katie Birchall, Laurel Kolakaluri, Deborah Baines & Anila Sukumaran

114. Nottingham University Hospital, Nottingham, UK

Megan Meredith, Lucy Morris, Lucy Ryan, Amy Clark, Julia Sampson, Cecilia Peters, Martin Dent, Margaret Langley, Saima Ashraf, Shuying Wei & Angela Andrew

115. Pilgrim Hospital, Lincoln, UK

Manish Chablani, Amy Kirkby & Kimberley Netherton

116. Queen Elizabeth Hospital, Birmingham, UK

Michelle Bates, Jo Dasgin, Jaspreet Gill, Annette Nilsson & James Scriven

117. Salford Royal Hospital, Manchester, UK

Elena Apetri, Cathrine Basikolo, Bethan Blackledge, Laura Catlow, Bethan Charles, Paul Dark, Reece Doonan, Jade Harris, Alice Harvey, Daniel Horner, Karen Knowles, Stephanie Lee, Diane Lomas, Chloe Lyons, Tracy Marsden, Danielle McLaughlan, Liam McMorrow, Jessica Pendlebury, Jane Perez, Maria Poulaka, Nicola Proudfoot, Melanie Slaughter, Kathryn Slevin, Melanie Taylor, Vicky Thomas, Danielle Walker, Angiy Michael & Matthew Collis

118. Tameside General Hospital, Ashton-under-Lyne, UK

Martyn Clark, Martina Coulding, Edward Jude, Jacqueline McCormick, Oliver Mercer, Darsh Potla, Hafiz Rehman, Heather Savill & Victoria Turner

119. The Tunbridge Wells Hospital and Maidstone Hospital, Maidstone, UK

Miriam Davey, David Golden & Rebecca Seaman

120. The Royal Oldham Hospital, Manchester, UK

Andrew Drummond, Jodie Hunt, Joy Dearden, Emma Dobson, Michelle Mulcahy, Sheila Munt, Grainne O'Connor, Jennifer Philbin, Chloe Rishton, Redmond Tully & Sarah Winnard

121. The Royal Papworth Hospital, Cambridge, UK

Lenka Cagova, Adama Fofano, Lucie Garner, Helen Holcombe, Sue Mepham, Alice Michael Mitchell, Lucy Mwaura, K. Praman, Alain Vuylsteke & Julie Zamikula

122. University College Hospital, London, UK

Georgia Bercades, David Brealey, Ingrid Hass, Niall MacCallum, Gladys Martir, Eamon Raith, Anna Reyes & Deborah Smyth

123. Withybush General Hospital, Haverfordwest, UK

Abigail Taylor, Rachel Anne Hughes, Helen Thomas, Alun Rees, Michaela Duskova, Janet Phipps, Suzanne Brooks & Michelle Edwards

124. Wythenshawe Hospital, Manchester, UK

Peter Alexander, Schvearn Allen, Joanne Bradley-Potts, Craig Brantwood, Jasmine Egan, Timothy Felton, Grace Padden, Luke Ward, Stuart Moss & Susannah Glasgow

125. Yeovil Hospital, Yeovil, UK

Kate Beesley, Sarah Board, Agnieszka Kubisz-Pudelko, Alison Lewis, Jess Perry, Lucy Pippard, Di Wood & Clare Buckley

126. Cumberland Infirmary, Carlisle, UK

Alison Brown, Jane Gregory, Susan O'Connell & Tim Smith

127. Darent Valley Hospital, Dartford, UK

Zakaula Belagodu, Bridget Fuller, Anca Gherman, Olumide Olufuwa, Remi Paramsothy, Carmel Stuart, Naomi Oakley, Charlotte Kamundi, David Tyl, Katy Collins, Pedro Silva, June Taylor, Laura King, Charlotte Coates, Maria Crowley, Phillipa Wakefield, Jane Beadle, Laura Johnson, Janet Sargeant & Madeleine Anderson

128. Dumfries and Galloway Royal Infirmary, Dumfries, UK

Catherine Jardine & Dewi Williams

129. Ealing Hospital, London, UK

Victoria Parris, Sheena Quaid & Ekaterina Watson

130. Fairfield General Hospital, Bury, UK

Julie Melville, Jay Naisbitt, Rosane Joseph, Maria Lazo, Olivia Walton & Alan Neal

131. George Eliot Hospital NHS Trust, Nuneaton, UK

Michaela Hill, Thogulava Kannan & Laura Wild

132. Glan Clwyd Hospital, Bodelwyddan, UK

Elizabeth Allan, Kate Darlington, Ffyon Davies, Jack Easton, Sumit Kumar, Richard Lean, Daniel Menzies, Richard Pugh, Xinyi Qiu, Llinos Davies, Hannah Williams, Jeremy Scanlon, Gwyneth Davies, Callum Mackay, Joanne Lewis & Stephanie Rees

133. Glangwili General Hospital, Camarthen, UK

Samantha Coetzee, Alistair Gales, Igor Otahal, Meena Raj & Craig Sell

134. The Great Western Hospital, Swindon, UK

Helen Langton, Rachel Prout, Malcolm Watters & Catherine Novis

135. Guys and St Thomas' Hospital, London, UK

Gill Arbane, Aneta Bociek, Sara Campos, Neus Grau, Tim Owen Jones, Rosario Lim, Martina Marotti, Marlies Ostermann, Manu Shankar-Hari & Christopher Whitton

136. Harefield Hospital, London, UK

Anthony Barron, Ciara Collins, Sundeep Kaul, Heather Passmore, Claire Prendergast, Anna Reed, Paula Rogers, Rajvinder Shokkar, Meriel Woodruff, Hayley Middleton, Oliver Polgar, Claire Nolan, Vicky Thwaites & Kanta Mahay

137. Harrogate and District NHS Foundation Trust, Harrogate, UK

Chunda Sri-Chandana, Joslan Scherewode, Lorraine Stephenson & Sarah Marsh

138. James Paget University Hospital NHS Trust, Great Yarmouth, UK

Amanda Ayers, Wendy Harrison & Julie North

139. King's Mill Hospital, Nottingham, UK

Mandy Gill, Paul Paul, Valli Ratnam, Sarah Shelton & Inez Wynter

140. Kingston Hospital, Kingston, UK

David Baptista, Rebecca Crowe, Rita Fernandes, Rosaleen Herdman-Grant, Anna Joseph, Adam Loveridge, India McKenley, Eriko Morino, Andres Naranjo, Richard Simms, Kathryn Sollest, Andrew Swain, Harish Venkatesh, Jacyntha Khera & Jonathan Fox

141. Lincoln County Hospital, Lincoln, UK

Russell Barber, Claire Hewitt, Annette Hilldrith, Karen Jackson-Lawrence, Sarah Shepardson, Maryanne Wills, Susan Butler, Silvia Tavares, Amy Cunningham, Julia Hindale & Sarwat Arif

142. Liverpool Heart and Chest Hospital, Liverpool, UK

Linsha George, Sophie Twiss & David Wright

143. Macclesfield District General Hospital, Macclesfield, UK

Maureen Holland, Natalie Keenan, Marc Lyons, Helen Wassall, Chris Marsh, Mervin Mahenthran, Emma Carter & Thomas Kong

144. Medway Maritime Hospital, Gillingham, UK

Oluronke Adanini, Nikhil Bhatia & Maines Msiska

145. Milton Keynes University Hospital, Milton Keynes, UK

Louise Mew, Esther Mwaura, Richard Stewart, Felicity Williams, Lynn Wren & Sara-Beth Sutherland

146. Morriston Hospital, Swansea, UK

Ceri Battle, Elaine Brinkworth, Rachel Harford, Carl Murphy, Luke Newey, Tabitha Rees, Marie Williams & Sophie Arnold

147. National Hospital for Neurology and Neurosurgery, London, UK

David Brealey, John Hardy, Henry Houlden, Eleanor Moncur, Eamon Raith, Ambreen Tariq & Arianna Tucci

148. Norfolk and Norwich University hospital (NNUH), Norwich, UK

Karen Convery, Deirdre Fottrell-Gould, Lisa Hudig, Jocelyn Keshet-Price, Georgina Randell & Katie Stammers

149. North Middlesex University Hospital NHS Trust, London, UK

Marwa Abdelrazik, Dhanalakshmi Bakthavatsalam, Munzir Elhassan, Arunkumar Ganesan, Anne Haldeos, Jeronimo Moreno-Cuesta, Dharam Purohit, Rachel Vincent, Kugan Xavier, Kumar Rohit, Frater Alasdair, Malik Saleem, Carter David, Samuel Jenkins, Zoe Lamond & Alanna Wall

150. Northumbria Healthcare NHS Foundation Trust, North Shields, UK

Bryan Yates, Jessica Reynolds, Helen Campbell, Maria Thompsom, Steve Dodds & Stacey Duffy

151. Peterborough City Hospital, Peterborough, UK

Deborah Butcher, Susie O' Sullivan & Nicola Butterworth-Cowin

152. Prince Charles Hospital, Merthyr Tydfil, UK

Bethan Deacon, Meg Hibbert, Carla Pothecary, Dariusz Tetla, Christopher Woodford, Latha Durga & Gareth Kennard-Holden

153. Royal Sussex County Hospital, Brighton, UK

Laura Ortiz-Ruiz de Gordo, Emily Peasgood & Claire Phillips

154. Princess Royal Hospital, Haywards Heath, UK

Denise Skinner, Jane Gaylard, Dee Mullan & Julie Newman

155. Princess of Wales Hospital, Llantrisant, UK

Ellie Davies, Lisa Roche & Sonia Sathe

156. Queen Alexandra Hospital, Portsmouth, UK

Lutece Brimfield, Zoe Daly, David Pogson & Steve Rose

157. Queen Elizabeth Hospital, London, UK

Amy Collins, Waqas Khaliq & Estefania Treus Gude

158. Queen Victoria Hospital, East Grinstead, UK

Julian Giles & Simon Booth

159. Queen's Hospital Burton, Burton-On-Trent, UK

Gillian Bell, Katy English, Amro Katary & Louise Wilcox

160. Raigmore Hospital, Inverness, UK

Rachael Campbell, Noreen Clarke, Jonathan Whiteside, Mairi Mascarenhas, Avril Donaldson, Joanna Matheson, Fiona Barrett, Marianne O’Hara, Laura O’Keefe & Clare Bradley

161. Rotherham General Hospital, Rotherham, UK

Dawn Collier, Anil Hormis, Rachel Walker & Victoria Maynard

162. Royal Blackburn Teaching Hospital, Blackburn, UK

Tahera Patel, Matthew Smith, Srikanth Chukkambotla, Aayesha Kazi, Janice Hartley, Joseph Dykes, Muhammad Hijazi, Sarah Keith, Meherunnisa Khan, Janet Ryan-Smith, Philippa Springle, Jacqueline Thomas, Nick Truman, Samuel Saad, Dabheoc Coleman, Christopher Fine, Roseanna Matt, Bethan Gay, Jack Dalziel, Syamlan Ali, Drew Goodchild, Rhiannan Harling, Ravi Bhatterjee, Wendy Goddard, Chloe Davison, Stephen Duberly, Jeanette Hargreaves & Rachel Bolton

163. Royal Preston Hospital, Preston, UK

Shondipon Laha, Mark Verlander & Alexandra Williams

164. Royal Surrey County Hospital, Guildford, UK

Helen Blackman, Ben Creagh-Brown, Sinead Donlon, Natalia Michalak-Glinska, Sheila Mtuwa, Veronika Pristopan, Armorel Salberg, Eleanor Smith, Sarah Stone, Charles Piercy, Jerik Verula, Dorota Burda, Rugia Montaser, Lesley Harden, Irving Mayangao, Cheryl Marriott, Paul Bradley & Celia Harris

165. Royal Albert Edward Infirmary, Wigan, UK

Joshua Cooper, Cheryl Finch, Sarah Liderth, Alison Quinn & Natalia Waddington

166. The Royal Alexandra Children’s Hospital, Brighton, UK

Katy Fidler, Emma Tagliavini & Kevin Donnelly

167. Royal Alexandra Hospital, Paisley, UK

Lynn Abel, Michael Brett, Brian Digby, Lisa Gemmell, James Hornsby, Patrick MacGoey, Pauline O'Neil, Richard Price, Natalie Rodden, Kevin Rooney, Radha Sundaram & Nicola Thomson

168. Royal Bolton Hospital, Bolton, UK

Rebecca Flanagan, Gareth Hughes, Scott Latham, Emma McKenna, Jennifer Anderson, Robert Hull & Kat Rhead

169. University Hospitals Dorset NHS Foundation Trust, Dorchester, UK

Debbie Branney, Jordan Frankham, Sally Pitts & Nigel White

170. Royal Brompton Hospital, London, UK

Daniele Cristiano, Natalie Dormand, Zohreh Farzad, Mahitha Gummadi, Kamal Liyanage, Sara Salmi, Geraldine Sloane, Vicky Thwaites, Mathew Varghese & Anelise C. Zborowski

171. Imperial College London, London, UK

Brijesh V. Patel

172. Royal Cornwall Hospital, Truro, UK

Sarah Bean, Karen Burt & Michael Spivey

173. Royal Free Hospital, London, UK

Christine Eastgate-Jackson, Helder Filipe, Daniel Martin, Amitaa Maharajh, Sara Mingo Garcia & Mark De Neef

174. Royal Glamorgan Hospital, Pontyclun, UK

Bethan Deacon, Ceri Lynch, Carla Pothecary, Lisa Roche, Gwenllian Sera Howe, Jayaprakash Singh, Keri Turner, Hannah Ellis & Natalie Stroud

175. Royal Gwent Hospital, Newport, UK

Shiney Cherian, Sean Cutler, Anne Emma Heron, Anna Roynon-Reed, Tamas Szakmany, Gemma Williams, Owen Richards & Yusuf Cheema

176. Northern General Hospital, Sheffield, UK

Norfaizan Ahmad, Joann Barker, Kris Bauchmuller, Sarah Bird, Kay Cawthon, Kate Harrington, Yvonne Jackson, Faith Kibutu, Becky Lenagh, Shamiso Masuko, Gary H. Mills, Ajay Raithatha, Matthew Wiles, Jayne Willson, Helen Newell, Alison Lye, Lorenza Nwafor, Claire Jarman, Sarah Rowland-Jones, David Foote, Joby Cole, Roger Thompson, James Watson, Lisa Hesseldon, Irene Macharia, Luke Chetam, Jacqui Smith, Amber Ford, Samantha Anderson, Kathryn Birchall, Kay Housley, Sara Walker, Leanne Milner, Helena Hanratty, Helen Trower, Patrick Phillips, Simon Oxspring & Ben Donne

177. Royal Hampshire County Hospital, Winchester, UK

Emily Bevan, Jane Martin, Dawn Trodd, Geoff Watson & Caroline Wrey Brown

178. Royal Manchester Children's Hospital, Manchester, UK

Lara Bunni, Claire Jennings, Monica Latif, Rebecca Marshall & Gayathri Subramanian

179. Royal Stoke University Hospital, Stoke-on-Trent, UK

Nageswar Bandla, Minnie GellamUCHO, Michelle Davies & Christopher Thompson

180. Salisbury District Hospital, Salisbury, UK

Phil Donnison, Fiona Trim & Beena Eapen

181. Sandwell General Hospital, Birmingham, UK

Cecilia Ahmed, Balvinder Baines, Sarah Clamp, Julie Colley, Risna Haq, Anne Hayes, Jonathan Hulme, Samia Hussain, Sibet Joseph, Rita Kumar, Zahira Maqsood & Manjit Purewal

182. Scarborough General Hospital, Scarborough, UK

Ben Chandler, Kerry Elliott, Janine Mallinson & Alison Turnbull

183. Scunthorpe General Hospital, Scunthorpe, UK

Kathy Dent, Elizabeth Horsley, Muhammad Nauman Akhtar, Sandra Pearson, Dorota Potoczna & Sue Spencer

184. Southmead Hospital, Bristol, UK

Hayley Blakemore, Borislava Borislavova, Beverley Faulkner, Emma Gendall, Elizabeth Goff, Kati Hayes, Matt Thomas, Ruth Worner, Kerry Smith & Deanna Stephens

185. St George's Hospital, London, UK

Carlos Castro Delgado, Deborah Dawson, Lijun Ding, Georgia Durrant, Obiageri Ezeobu, Sarah Farnell-Ward, Abiola Harrison, Rebecca Kanu, Susannah Leaver, Elena Maccacari, Soumendu Manna, Romina Pepermans Saluzzio, Joana Queiroz, Tinashe Samakomva, Christine Sicat, Joana Texeira, Edna Fernandes Da Gloria, Ana Lisboa, John Rawlins, Jisha Mathew, Ashley Kinch, William James Hurt, Nirav Shah, Victoria Clark, Maria Thanasi, Nikki Yun & Kamal Patel

186. St Mary's Hospital, Newport, UK

Vikki Crickmore, Gabor Debreceni, Joy Wilkins & Liz Nicol

187. Stoke Mandeville Hospital, Aylesbury, UK

Iona Burn, Geraldine Hambrook, Katarina Manso, Ruth Penn, Pradeep Shanmugasundaram, Julie Tebbutt & Danielle Thornton

188. Sunderland Royal Hospital, Sunderland, UK

Anthony Rostron, Alistair Roy, Lindsey Woods, Sarah Cornell, Fiona Wakinshaw, Kimberley Rogerson & Jordan Jarmain

189. Alexandra Hospital, Redditch and Worcester Royal Hospital, Worcester, UK

Peter Anderson, Katie Archer, Karen Austin, Caroline Davis, Alison Durie, Olivia Kelsall, Jessica Thrush, Charlie Vigurs, Laura Wild, Hannah-Louise Wood, Helen Tranter, Alison Harrison, Nicholas Cowley, Michael McAlindon, Andrew Burtenshaw, Stephen Digby, Emma Low, Aled Morgan, Naiara Cother, Tobias Rankin, Sarah Clayton & Alex McCurdy

190. The Christie NHS Foundation Trust, Manchester, UK

Suzanne Allibone, Roman Mary-Genetu, Vidya Kasipandian, Amit Patel, Ainho Mac, Anthony Murphy, Parisa Mahjoob, Roonak Nazari, Lucy Worsley & Andrew Fagan

191. The Queen Elizabeth Hospital, King's Lynn, UK

Inthakab Ali Mohamed Ali, Karen Beaumont, Mark Blunt, Zoe Coton, Hollie Curgenven, Mohamed Elsaadany, Kay Fernandes, Sameena Mohamed Ally, Harini Rangarajan, Varun Sarathy, Sivarupan Selvanayagam, Dave Vedage & Matthew White

192. The Royal Liverpool University Hospital, Liverpool, UK

Jaime Fernandez-Roman, David O. Hamilton, Emily Johnson, Brian Johnston, Maria Lopez Martinez, Suleman Mulla, David Shaw, Alicia A. C. Waite, Victoria Waugh, Ingeborg D. Welters & Karen Williams

193. The Royal Marsden NHS Foundation Trust, London, UK

Thomas Bemand, Ethel Black, Arnold Dela Rosa, Ryan Howle, Shaman Jhanji, Ravishankar Rao Baikady, Kate Colette Tatham & Benjamin Thomas

194. Torbay Hospital, Torquay, UK

Matthew Halkes, Pauline Mercer & Lorraine Thornton

195. University Hospital Monklands, Airdrie, UK

Joe West, Tracy Baird & Jim Ruddy

196. University Hospital Lewisham, London, UK

Waqas Khaliq & Rosie Reece-Anthony

197. University Hospital North Durham, Darlington, UK

Mark Birt, Amanda Cowton, Andrea Kay, Melanie Kent, Kathryn Potts, Ami Wilkinson, Suzanne Naylor & Ellen Brown

198. University Hospital of North Tees, Stockton-on-Tees, UK

Michele Clark & Sarah Purvis

199. University Hospital of Wales, Cardiff, UK

Jade Cole, Michelle Davies, Rhys Davies, Donna Duffin, Helen Hill, Ben Player, Emma Thomas & Angharad Williams

200. University Hospital Wishaw, Wishaw, UK

Claire Marie Beith, Karen Black, Suzanne Clements, Alan Morrison, Dominic Strachan, Margaret Taylor, Michelle Clarkson, Stuart D'Sylva & Kathryn Norman

201. Victoria Hospital, Kirkcaldy, UK

Tina Coventry, Susan Fowler, Michael MacMahon & Amanda McGregor

202. Warrington General Hospital, Warrington, UK

Ailbhe Brady, Rebekah Chan, Jeff Little, Shane McIvor, Helena Prady, Helen Whittle & Bijoy Mathew

203. West Cumberland Hospital, Whitehaven, UK

Melanie Clapham, Rosemary Harper, Una Poultney, Polly Rice, Tim Smith & Rachel Mutch

204. Western Sussex Hospitals, Chichester, UK

Yolanda Baird, Aaron Butler, Indra Chadbourn, Linda Folkes, Heather Fox, Amy Gardner, Raquel Gomez, Gillian Hobden, Luke Hodgson, Kirsten King, Michael Margarson, Tim Martindale, Emma Meadows, Dana Raynard, Yvette Thirlwall, David Helm & Jordi Margalef

205. Whiston Hospital, Prescot, UK

Sandra Greer, Karen Shuker & Ascanio Tridente

206. York Hospital, York, UK

Isobel Birkinshaw, Joseph Carter, Kate Howard, Joanne Ingham, Rosie Joy, Harriet Pearson, Samantha Roche & Zoe Scott

207. Ysbyty Gwynedd, Bangor, UK

Ellen Knights, Alicia Price, Alice Thomas & Chris Thorpe

208. Countess of Chester Hospital, Chester, UK

Azmerelda Abraheem, Peter Bamford, Kathryn Cawley, Charlie Dunmore, Maria Faulkner, Rumanah Girach, Helen Jeffrey, Rhianna Jones, Emily London, Imrun Nagra, Farah Nasir, Hannah Sainsbury & Clare Smedley

209. Croydon University Hospital, Croydon, UK

Reena Khade, Ashok Sundar & George Tsinaslanidis

210. Diana Princess of Wales Hospital, Grimsby, UK

Teresa Behan, Caroline Burnett, Jonathan Hatton, Elaine Heeney, Atideb Mitra, Maria Newton, Rachel Pollard & Rachael Stead

211. Dorset County Hospital, Dorchester, UK

Jenny Birch, Laura Bough, Josie Goodsell, Rebecca Tutton, Patricia Williams, Sarah Williams & Barbara Winter-Goodwin

212. Forth Valley Royal Hospital, Falkirk, UK

Fiona Auld, Joanne Donnachie, Ian Edmond, Lynn Prentice, Nikole Runciman, Dario Salutous, Lesley Symon, Anne Todd, Patricia Turner, Abigail Short, Laura Sweeney, Euan Murdoch & Dhaneesha Senaratne

213. Furness General Hospital, Barrow-in-Furness, UK

Karen Burns & Andrew Higham

214. Alder Hey Children's Hospital, Liverpool, UK

Taya Anderson, Dan Hawcutt, Laura O' Malley, Laura Rad, Naomi Rogers, Paula Saunderson, Kathryn Sian Allison, Deborah Afolabi, Jennifer Whitbread, Dawn Jones & Rachael Dore

215. Derriford Hospital, Plymouth, UK

Liana Lankester, Nikitas Nikitas, Colin Wells, Bethan Stowe & Kayleigh Spencer

216. Glasgow Royal Infirmary, Glasgow, UK

Susanne Cathcart, Katharine Duffy, Alex Puxty, Kathryn Puxty, Lynne Turner, Jane Ireland & Gary Semple

217. Glenfield Hospital, Leicester, UK

Peter Barry

218. Gloucestershire Royal Hospital, Gloucester, UK

Paula Hilltout, Jayne Evitts, Amanda Tyler & Joanne Waldron

219. Golden Jubilee National Hospital, Clydebank, UK

Val Irvine & Benjamin Shelley

220. Great Ormond St Hospital and UCL Great Ormond St Institute of Child Health NIHR Biomedical Research Centre, London, UK

Olugbenga Akinkugbe, Alasdair Bamford, Emily Beech, Holly Belfield, Michael Bell, Charlene Davies, Gareth A. L. Jones, Tara McHugh, Hamza Meghari, Lauran O'Neill, Mark J. Peters, Samiran Ray & Ana Luisa Tomas

221. Homerton University Hospital Foundation NHS Trust, London, UK

Amy Easthope, Claire Gorman, Abhinav Gupta, Elizabeth Timlick & Rebecca Brady

222. James Cook University Hospital, Middlesbrough, UK

Stephen Bonner, Keith Hugill, Jessica Jones & Steven Liggett

223. John Radcliffe Hospital, Oxford, UK

Archana Bashyal, Neil Davidson, Paula Hutton, Stuart McKechnie & Jean Wilson

224. Leicester Royal Infirmary, Leicester, UK

Neil Flint, Patel Rekha & Dawn Hales

225. Lister Hospital, Stevenage, UK

Carina Cruz & Natalie Pattison

226. New Cross Hospital, Wolverhampton, UK

Shameer Gopal, Nichola Harris, Victoria Lake, Stella Metherell & Elizabeth Radford

227. Royal Victoria Infirmary, Newcastle Upon Tyne, UK

Ian Clement, Bijal Patel, A. Gulati, Carole Hays, K. Webster, Anne Hudson, Andrea Webster, Elaine Stephenson, Louise McCormack, Victoria Slater, Rachel Nixon, Helen Hanson, Maggie Fearby, Sinead Kelly, Victoria Bridgett & Philip Robinson

228. Ninewells Hospital, Dundee, UK

Christine Almaden-Boyle, Pauline Austin, Louise Cabrelli, Stephen Cole, Matt Casey, Susan Chapman & Clare Whyte

229. North Devon District Hospital, Barnstaple, UK

Adam Brayne, Emma Fisher, Jane Hunt, Peter Jackson, Duncan Kaye, Nicholas Love, Juliet Parkin, Victoria Tuckey, Lynne van Koutrik, Sasha Carter, Benedict Andrew, Louise Findlay & Katie Adams

230. North Manchester General Hospital, Manchester, UK

Michelle Bruce, Karen Connolly, Tracy Duncan, Helen T.-Michael, Gabriella Lindergard, Samuel Hey, Claire Fox, Jordan Alfonso, Laura Jayne Durrans, Jacinta Guerin, Bethan Blackledge, Jade Harris, Martin Hruska, Ayaa Eltayeb, Thomas Lamb, Tracey Hodgkiss, Lisa Cooper & Joanne Rothwell

231. Northwick Park Hospital, London, UK

Catherine Dennis, Alastair McGregor, Victoria Parris, Sinduya Srikanan & Anisha Sukha

232. Prince Philip Hospital, Lianelli, UK

Kim Davies, Linda O' Brien, Zohra Omar, Igor Otahal, Emma Perkins, Tracy Lewis & Isobel Sutherland

233. Pinderfields General Hospital, Wakefield, UK

Hollie Brooke, Sarah Buckley, Jose Cebrian Suarez, Ruth Charlesworth, Karen Hansson, John Norris, Alice Poole, Alastair Rose, Rajdeep Sandhu, Brendan Sloan, Elizabeth Smithson, Muthu Thirumaran, Veronica Wagstaff & Alexandra Metcalfe

234. Poole Hospital, Poole, UK

Julie Camsooksai, Charlotte Humphrey, Sarah Jenkins, Henrik Reschreiter, Beverley Wadams & Yasmin DeAth

235. Royal Shrewsbury Hospital, Shrewsbury, UK

Colene Adams, Anita Agasou, Amy Bowes, Pauline Boyle, Mandy Carnahan, Anne Carter, Danielle Childs, Kelly Hard, Yasmin Hussain, Michael Leigh, Rachel Rikunenko, Jo Stickley, Helen Tivenan & Rebecca Wilcox

236. Princess Royal Hospital, Telford, UK

Tracie Arden, Mandy Beekes, Heather Button, Nigel Capps, Denise Donaldson, Fran Hurford, Ayesha Javaid, James Jones, Sanal Jose, Terry Martin, Helen Millward, Nichola Motherwell, Julie Summers, Louise Ting & Louise Tonks

237. Queen Elizabeth Hospital Gateshead, Gateshead, UK

Maria Bokhari, Vanessa Linnett, Rachael Lucas, Wendy McCormick, Jenny Ritzema, Amanda Sanderson & Helen Wild

238. Queen Elizabeth University Hospital, Glasgow, UK

Nicola Baxter, Steven Henderson, Sophie Kennedy-Hay, Christopher McParland, Laura Rooney, Malcolm Sim & Gordan McCreath

239. Royal Berkshire NHS Foundation Trust, Reading, UK

Mark Brunton, Jess Caterson, Holly Coles, Matthew Frise, Sabi Gurung Rai, Nicola Jacques, Liza Keating, Emma Tilney, Shauna Bartley & Parminder Bhui

240. Royal Derby Hospital, Derby, UK

Charlotte Downes, Kathleen Holding, Katie Riches, Mary Hilton, Mel Hayman, Deepak Subramanian & Priya Daniel

241. Royal Devon and Exeter Hospital, Exeter, UK

Letizia Zitter, Sarah Benyon, Suzie Marriott, Linda Park, Samantha Keenan, Elizabeth Gordon, Helen Quinn & Kizzy Baines

242. Royal Infirmary of Edinburgh, Edinburgh, UK

Gillian Andrew, J. Kenneth Baillie, Lucy Barclay, Marie Callaghan, Rachael Campbell, Sarah Clark, Dave Hope, Lucy Marshall, Corrienne McCulloch, Kate Briton, Jo Singleton & Sophie Birch

243. Royal Lancaster Infirmary, Lancaster, UK

Andrew Higham, Andrew Higham, Kerry Simpson & Jayne Craig

244. Royal United Hospital, Bath, UK

Carrie Demetriou, Charlotte Eckbad, Sarah Hierons, Lucy Howie, Sarah Mitchard, Lidia Ramos, Alfredo Serrano-Ruiz, Katie White & Fiona Kelly

245. Russells Hall Hospital, Dudley, UK

Vishal Amin, Elena Anastasescu, Vikram Anumakonda, Komala Karthik, Rizwana Kausar, Karen Reid, Jacqueline Smith & Janet Imeson-Wood

246. Sheffield Children's Hospital, Sheffield, UK

Arianna Bellini, Jade Bryant, Anton Mayer, Amy Pickard, Nicholas Roe, Jason Sowter & Alex Howlett

247. Southampton General Hospital, Southampton, UK

Kristine Criste, Rebecca Cusack, Kim Golder, Hannah Golding, Oliver Jones, Samantha Leggett, Michelle Male, Martyna Marani, Kirsty Prager, Toran Williams, Belinda Roberts & Karen Salmon

248. Southend University Hospital, Westcliff-on-Sea, UK

Prisca Gondo, B. Hadebe, Abdul Kayani & Bridgett Masunda

249. Southport and Formby District General Hospital, Ormskirk, UK

Ashar Ahmed, Anna Morris & Srinivas Jakkula

250. St James's University Hospital and Leeds General Infirmary, Leeds, UK

Kate Long, Simon Whiteley, Elizabeth Wilby & Bethan Ogg

251. Bristol Royal Infirmary, Bristol, UK

Jeremy Bewley, Zoe Garland, Lisa Grimmer, Bethany Gumbrill, Rebekah Johnson, Katie Sweet, Denise Webster & Georgia Efford

252. Stepping Hill Hospital, Stockport, UK

Sara Bennett, Emma Goodwin, Matthew Jackson, Alissa Kent, Clare Tibke, Wiesia Woodyatt & Ahmed Zaki

253. The Princess Alexandra Hospital, Harlow, UK

Amelia Daniel, Joanne Finn, Rajnish Saha, Nikki Staines & Amy Easthope

254. University Hospital Crosshouse, Kilmarnock, UK

J. Allan, T. Geary, Gordon Houston, A. Meikle & P. O'Brien

255. University Hospital Hairmyres, East Kilbride, UK

Dina Bell, Rosalind Boyle, Katie Douglas, Lynn Glass, Emma Lee, Liz Lennon & Austin Rattray

256. Craigavon Area Hospital, Craigavon, UK

Rob Charnock, Denise McFarland & Denise Cosgrove

257. Warwick Hospital, Warwick, UK

Ben Attwood & Penny Parsons

258. West Middlesex Hospital, Isleworth, UK

Metod Oblak, Monica Popescu & Mini Thankachen

259. Whittington Hospital, London, UK

Abdelhakim Altabaibeh, Ana Alvaro, Kayleigh Gilbert, Louise Ma, Loreta Mostoles, Chetan Parmar, Kathryn Simpson, Champa Jetha, Lauren Booker & Anezka Pratley

260. William Harvey Hospital, Ashford, UK

Tracey Cosier, Gemma Millen, Neil Richardson, Natasha Schumacher, Heather Weston & James Rand

261. Section of Molecular Virology, Imperial College London, London, UK

Wendy S. Barclay

262. Antimicrobial Resistance and Hospital Acquired Infection Department, Public Health England, London, UK

Meera Chand

263. Department of Infectious Disease, Imperial College London, London, UK

Graham S. Cooke & Shiranee Sriskandan

264. National Infection Service, Public Health England, London, UK

Jake Dunning & Maria Zambon

265. MRC-University of Glasgow Centre for Virus Research, Glasgow, UK

Ana da Silva Filipe, Massimo Palmarini, David L. Robertson, Janet T. Scott, Emma C. Thomson & Sarah E. McDonald

266. Liverpool School of Tropical Medicine, Liverpool, UK

Tom Fletcher

267. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

Julian A. Hiscox

268. Virology Reference Department, National Infection Service, Public Health England, London, UK

Samreen Ijaz

269. Department of Pharmacology, University of Liverpool, Liverpool, UK

Saye Khoo

270. Nuffield Department of Medicine, University of Oxford, Oxford, UK

Paul Klenerman

271. Nottingham University Hospitals NHS Trust, Nottingham, UK

Wei Shen Lim

272. Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK

Alexander J. Mentzer

273. ISARIC Global Support Centre, Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Laura Merson, Louise Sigfrid & Gail Carson

274. Division of Infection and Immunity, University College London, London, UK

Mahdad Noursadeghi

275. Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK

Shona C. Moore, William A. Paxton & Georgios Pollakis

276. Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London, London, UK

Nicholas Price

277. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Andrew Rambaut

278. Department of Pediatrics and Virology, Imperial College London, London, UK

Vanessa Sancho-Shimizu

279. The Florey Institute for Host-Pathogen Interactions, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

Thushan de Silva

280. Division of Structural Biology, The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

David Stuart

281. Blood Borne Virus Unit, Virus Reference Department, National Infection Service, Public Health England, London, UK

Richard S. Tedder

282. Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

A. A. Roger Thompson

283. Institute for Global Health, University College London, London, UK

Rishi K. Gupta

284. Molecular and Clinical Cancer Medicine, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK

Carlo Palmieri

285. Department of Child Life and Health, University of Edinburgh, Edinburgh, UK

Olivia V. Swann

286. Section of Biomolecular Medicine, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

Marc-Emmanuel Dumas, Julian L. Griffin, Zoltan Takats, Petros Andrikopoulos, Anthonia Osagie, Michael Olanipekun & Sonia Liggi

287. Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK

Kanta Chechi

288. National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

Matthew R. Lewis, Gonçalo dos Santos Correia, Caroline J. Sands, Panteleimon Takis & Lynn Maslen

289. Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK

William Greenhalf

290. Institute of Translational Medicine, University of Liverpool, Liverpool, UK

Victoria Shaw

291. Intensive Care Unit, Royal Infirmary Edinburgh, Edinburgh, UK

Seán Keating

292. University of Liverpool, Liverpool, UK

Sarah Cole, Katie A. Ahmed, Jane A. Armstrong, Milton Ashworth, Innocent G. Asiimwe, Siddharth Bakshi, Samantha L. Barlow, Laura Booth, Katie Bullock, Benjamin W. A. Catterall, Jordan J. Clark, Emily A. Clarke, Louise Cooper, Helen Cox, Christopher Davis, Oslem Dincarslan, Chris Dunn, Philip Dyer, Angela Elliott, Anthony Evans, Lorna Finch, Lewis W. S. Fisher, Terry Foster, Isabel Garcia-Dorival, William Greenhalf, Philip Gunning, Catherine Hartley, Rebecca L. Jensen, Christopher B. Jones, Trevor R. Jones, Shadia Khandaker, Katharine King, Robyn T.

Kiy, Chrysa Koukorava, Suzannah Lant, Diane Latawiec, Lara Lavelle-Langham, Lauren Lett, Lucia A. Livoti, Maria Mancini, Sarah McDonald, Laurence McEvoy, Soeren Metelmann, Nahida S. Miah, Joanna Middleton, Joyce Mitchell, Shona C. Moore, Ellen G. Murphy, Rebekah Penrice-Randal, Jack Pilgrim, Tessa Prince, Will Reynolds, P. Matthew Ridley, Debby Sales, Victoria E. Shaw, Rebecca K. Shears, Benjamin Small, Krishanthi S. Subramaniam, Agnieska Szemiel, Jolanta Tanianis-Hughes, Jordan Thomas, Erwan Trochu, Libby van Tonder, Eve Wilcock, J. Eunice Zhang, Lisa Flaherty, Nicole Maziere, Emily Cass, Alejandra Doce Carracedo, Nicola Carlucci, Anthony Holmes & Hannah Massey

293. University of Glasgow, Glasgow, UK

Benjamin Brennan, Annette Lake, Daniella Lefteri, John McLauchlan & Aislynn Taggart

294. Edinburgh Clinical Research Facility, Western General Hospital, University of Edinburgh, Edinburgh, UK

Richard Clark, Audrey Coutts, Lorna Donnelly, Angie Fawkes, Tammy Gilchrist, Katarzyna Hafezi, Louise Macgillivray, Alan Maclean, Sarah McCafferty, Kirstie Morrice, Lee Murphy & Nicola Wrobel

295. Department of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

Maaike Swets

296. Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Katherine Schon, Anita Furlong & Heather Biggs

297. Genotek, Moscow, Russia

Alexander Rakitko, Valery Ilinsky, Danat Yermakovich, Iaroslav Popov, Alexander Chernitsov, Elena Kovalenko, Anna

Krasnenko, Nikolay Plotnikov, Ivan Stetsenko & Anna Kim

298. Helix, San Mateo, CA, USA

Elizabeth T. Cirulli, Kelly M. Schiabor Barrett, Alexandre Bolze, Simon White, Nicole L. Washington, James T. Lu, Stephen Riffle, Francisco Tanudjaja, Xueqing Wang, Jimmy M. Ramirez III, Nicole Leonetti & Efren Sandoval

299. Center for Genomic Medicine, Desert Research Institute, Reno, NV, USA

Iva Neveux & Joseph J. Grzymski

300. Renown Health, Reno, NV, USA

Shaun Dabe

301. 24Genetics, Boston, MA, USA

Juan Ignacio Esteban Miñano, Manuel de la Mata Pazos, Luciano Cerrato & Lasse Folkersen

302. Hospital La Paz Institute for Health Research, Madrid, Spain

Luis A. Aguirre, Eduardo López-Collazo, Roberto Lozano-Rodríguez, José Avendaño-Ortiz, Verónica Terrón Arcos, Karla Marina Montalbán-Hernández, Jaime Valentín Quiroga, Alejandro Pascual-Iglesias, Charbel Maroun-Eid & Alejandro Martín-Quirós

303. Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan

Ho Namkoong, Koichi Fukunaga, Makoto Ishii, Hiroki Kabata, Katsunori Masaki, Hirofumi Kamata, Shinnosuke Ikemura, Shotaro Chubachi, Satoshi Okamori, Hideki Terai, Hiromu Tanaka, Atsuho Morita, Ho Lee & Takanori Asakura

304. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan

Yukinori Okada, Ryuya Edahiro, Kyuto Sonehara, Yuya Shirai & Ken Suzuki

305. Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan

Yukinori Okada

306. Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan

Yukinori Okada & Atsushi Kumanogoh

307. Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Seiya Imoto

308. Laboratory of Viral Infection I, Department of Infection Control and Immunology, Ōmura Satoshi Memorial Institute & Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan

Kazuhiko Katayama

309. Department of Surgery, Keio University School of Medicine, Tokyo, Japan

Yuko Kitagawa

310. Department of Organoid Medicine, Keio University School of Medicine, Tokyo, Japan

Toshiro Sato

311. Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan

Naoki Hasegawa

312. Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan

Atsushi Kumanogoh, Ryuya Edahiro, Yuya Shirai, Yoshito Takeda, Haruhiko Hirata, Takayuki Shiroyama, Yuichi Maeda, Takuro Nii, Yoshimi Noda, Takayuki Niitsu, Yuichi Adachi, Takatoshi Enomoto, Saori Amiya & Reina Hara

313. Department of Immunopathology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan

Atsushi Kumanogoh

314. Institute of Research, Tokyo Medical and Dental University, Tokyo, Japan

Akinori Kimura

315. Department of Insured Medical Care Management, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan

Masumi Ai

316. Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Chiba, Japan

Katsushi Tokunaga & Yosuke Omae

317. Division of Gastroenterology and Hepatology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan

Takanori Kanai, Kosaku Nanki & Yohei Mikami

318. M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan

Satoru Miyano, Kunihiko Takahashi, Tatsuhiko Anzai, Takanori Hasegawa & Satoshi Ito

319. Department of Pathology and Tumor Biology Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan

Seishi Ogawa

320. Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden

Seishi Ogawa

321. Department of Emergency and Critical Care Medicine, Keio University School of Medicine, Tokyo, Japan

Junichi Sasaki

322. Department of Anesthesiology, Keio University School of Medicine, Tokyo, Japan

Hiroshi Morisaki

323. Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan

Yoshifumi Uwamino

324. Division of Infection Control and Prevention, Osaka University Hospital, Suita, Japan

Kazunori Tomono

325. Department of Biomedical Ethics and Public Policy, Osaka University Graduate School of Medicine, Suita, Japan

Kazuto Kato

326. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan

Fumihiko Matsuda & Meiko Takahashi

327. Department of Pulmonary Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

Nobuyuki Hizawa

328. Department of Neurosurgery, Faculty of Medicine, The University of Tokyo, Tokyo, Japan

Satoru Miyawaki

329. Laboratory of Immune Regulation, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita, Japan

Yuichi Maeda & Takuro Nii

330. Medical Innovation Promotion Center, Tokyo Medical and Dental University, Tokyo, Japan

Ryuji Koike

331. Clinical Research Center, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan

Akifumi Endo

332. Department of Medical Informatics, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan

Yuji Uchimura

333. Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan

Yasunari Miyazaki, Takayuki Honda & Tomoya Tateishi

334. Clinical Laboratory, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan

Shuji Tohda, Naoya Ichimura, Kazunari Sonobe, Chihiro Sassa & Jun Nakajima

335. Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan

Yasuhiro Nannya

336. Department of Respiratory Medicine, Graduate School of Medicine, Faculty of Medicine, Juntendo University, Tokyo, Japan

Kazuhide Takahashi, Norihiro Harada, Haruhi Takagi & Ai Nakamura

337. Department of Emergency and Disaster Medicine, Graduate School of Medicine, Faculty of Medicine, Juntendo University, Tokyo, Japan

Makoto Hiki

338. Department of Cardiovascular Biology and Medicine, Graduate School of Medicine, Faculty of Medicine, Juntendo University, Tokyo, Japan

Makoto Hiki

339. Department of Respiratory Medicine, Tokyo Women's Medical University, Tokyo, Japan

Etsuko Tagaya & Ken Arimura

340. Department of General Medicine, Tokyo Women's Medical University, Tokyo, Japan

Masatoshi Kawana

341. Department of Respiratory Medicine, Saitama Cardiovascular and Respiratory Center, Saitama, Japan

Takashi Ishiguro, Noboru Takayanagi, Taisuke Isono, Yotaro Takaku & Kenji Takano

342. Kawasaki Municipal Ida Hospital, Kanagawa, Japan

Ryusuke Anan, Yukiko Nakajima, Yasushi Nakano & Kazumi Nishio

343. Saitama Medical Center, Internal Medicine, Japan Community Healthcare Organization (JCHO), Saitama, Japan

Soichiro Ueda & Reina Hayashi

344. Saitama City Hospital, Saitama, Japan

Hiroki Tateno, Isano Hase, Shuichi Yoshida & Shoji Suzuki

345. Division of Infection Control, Eiju General Hospital, Tokyo, Japan

Keiko Mitamura

346. Department of Pulmonary Medicine, Eiju General Hospital, Tokyo, Japan

Fumitake Saito

347. Department of Respiratory Medicine, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan

Tetsuya Ueda, Masanori Azuma, Tadao Nagasaki & Yoshinori Hasegawa

348. Division of Respirology, Rheumatology, and Neurology, Department of Internal Medicine Kurume University School of Medicine, Fukuoka, Japan

Masaki Okamoto

349. Department of Infection Control, Osaka Saiseikai Nakatsu Hospital,  
Osaka, Japan

Yoshinori Yasui

350. Department of Infectious Diseases, Tosei General Hospital, Aichi,  
Japan

Yoshikazu Mutoh

351. Fukujuji Hospital, Kiyose, Japan

Takashi Yoshiyama

352. Department of Emergency and Critical Care Medicine, Tokyo  
Women's Medical University Medical Center East, Tokyo, Japan

Tomohisa Shoko, Mitsuaki Kojima & Tomohiro Adachi

353. Department of Medicine, Tokyo Women's Medical University Medical  
Center East, Tokyo, Japan

Motonao Ishikawa

354. Department of Pediatrics, Tokyo Women's Medical University  
Medical Center East, Tokyo, Japan

Kenichiro Takahashi

355. Japan Community Healthcare Organization Kanazawa Hospital,  
Kanazawa, Japan

Kazuyoshi Watanabe

356. Division of Pulmonary Medicine, Department of Internal Medicine,  
Federation of National Public Service Personnel Mutual Aid  
Associations, Tachikawa Hospital, Tachikawa, Japan

Tadashi Manabe, Fumimaro Ito, Takahiro Fukui, Yohei Funatsu & Hidefumi Koh

357. Department of Respiratory Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan

Yoshihiro Hirai, Hidetoshi Kawashima & Atsuya Narita

358. Department of General Internal Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan

Kazuki Niwa & Yoshiyuki Sekikawa

359. Department of Emergency and Critical Care Medicine, Kansai Medical University General Medical Center, Kirakata, Japan

Fukuki Saito, Kazuhisa Yoshiya & Tomoyuki Yoshihara

360. Department of Respiratory Medicine, Kitasato University, Kitasato Institute Hospital, Tokyo, Japan

Yusuke Suzuki, Sohei Nakayama & Keita Masuzawa

361. Ishikawa Prefectural Central Hospital, Kanazawa, Japan

Koichi Nishi, Masaru Nishitsuji & Maiko Tani

362. Internal Medicine, Sano Kosei General Hospital, Sano, Japan

Takashi Inoue, Toshiyuki Hirano & Keigo Kobayashi

363. Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan

Naoki Miyazawa, Yasuhiro Kimura & Reiko Sado

364. Kanagawa Cardiovascular and Respiratory Center, Yokohama, Japan

Takashi Ogura, Hideya Kitamura & Kota Murohashi

365. Saiseikai Utsunomiya Hospital, Utsunomiya, Japan

Ichiro Nakachi, Rie Baba & Daisuke Arai

366. Department of Respiratory Medicine, KKR Sapporo Medical Center, Sapporo, Japan

Satoshi Fuke & Hiroshi Saito

367. Internal Medicine, Internal Medicine Center, Showa University Koto Toyosu Hospital, Tokyo, Japan

Naota Kuwahara, Akiko Fujiwara & Takenori Okada

368. Department of Respiratory Medicine, Toyohashi Municipal Hospital, Toyohashi, Japan

Tomoya Baba, Junya Noda & Shuko Mashimo

369. Keiyu Hospital, Yokohama, Japan

Kazuma Yagi, Tetsuya Shiomi & Mizuha Hashiguchi

370. Department of Rheumatology, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan

Toshio Odani

371. Department of Respiratory Medicine, National Hospital Organization Tokyo Medical Center, Tokyo, Japan

Takao Mochimaru & Yoshitaka Oyamada

372. Department of Allergy, National Hospital Organization Tokyo Medical Center, Tokyo, Japan

Takao Mochimaru & Yoshitaka Oyamada

373. Department of General Internal Medicine and Infectious Diseases,  
National Hospital Organization Tokyo Medical Center, Tokyo, Japan

Nobuaki Mori

374. Japanese Red Cross Musashino Hospital, Musashino, Japan

Namiki Izumi, Kaoru Nagata & Reiko Taki

375. Department of Respiratory Medicine, Tohoku University Graduate  
School of Medicine, Sendai, Japan

Koji Murakami, Mitsuhiro Yamada & Hisatoshi Sugiura

376. Division of Respiratory Medicine, Department of Internal Medicine,  
Nihon University School of Medicine, Tokyo, Japan

Kentaro Hayashi, Tetsuo Shimizu & Yasuhiro Gon

377. Department of Emergency and Critical Care Medicine, St Marianna  
University School of Medicine, Kawasaki, Japan

Shigeki Fujitani & Toru Yoshida

378. Division of General Internal Medicine, Department of Internal  
Medicine, St Marianna University School of Medicine, Kawasaki,  
Japan

Tomoya Tsuchida

379. National Hospital Organization Kanazawa Medical Center, Kanazawa,  
Japan

Takashi Kagaya, Toshiyuki Kita & Satoru Sakagami

380. Division of Infectious Diseases and Respiratory Medicine, Department  
of Internal Medicine, National Defense Medical College, Tokorozawa,  
Japan

Yoshifumi Kimizuka & Akihiko Kawana

381. Department of Emergency and Critical Care Medicine, Faculty of Medicine, Fukuoka University, Fukuoka, Japan

Yoshihiko Nakamura & Hiroyasu Ishikura

382. Department of Infection Control, Fukuoka University Hospital, Fukuoka, Japan

Tohru Takata

383. Tokyo Saiseikai Central Hospital, Tokyo, Japan

Takahide Kikuchi, Daisuke Taniyama & Morio Nakamura

384. Department of Internal Medicine, Fukuoka Tokushukai Hospital, Kasuga, Japan

Nobuhiro Kodama, Yasunari Kaneyama & Shunsuke Maeda

385. Department of Infectious Disease and Clinical Research Institute, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan

Yoji Nagasaki & Sayoko Ishihara

386. Department of Respirology, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan

Masaki Okamoto

387. Matsumoto City Hospital, Matsumoto, Japan

Akihiro Ito

388. Uji-Tokushukai Medical Center, Uji, Japan

Yusuke Chihara, Mayumi Takeuchi & Keisuke Onoi

389. Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan

Naozumi Hashimoto, Keiko Wakahara & Akira Ando

390. Department of Respiratory Medicine, Fujisawa City Hospital, Fujisawa, Japan

Makoto Masuda, Aya Wakabayashi & Hiroki Watanabe

391. Sapporo City General Hospital, Sapporo, Japan

Hisako Sageshima

392. Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, Chiba, Japan

Taka-Aki Nakada, Ryuzo Abe & Tadanaga Shimada

393. Division of Respiratory Medicine, Social Welfare Organization Saiseikai Imperial Gift Foundation, Saiseikai Kumamoto Hospital, Kumamoto, Japan

Kodai Kawamura, Kazuya Ichikado & Kenta Nishiyama

394. Department of Anesthesiology and Intensive Care Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan

Masaki Yamasaki & Satoru Hashimoto

395. Ome Municipal General Hospital, Ome, Japan

Yu Kusaka, Takehiko Ohba & Susumu Isogai

396. Hanwa Daini Hospital, Osaka, Japan

Minoru Takada & Hidenori Kanda

397. Department of Respiratory Internal Medicine, St Marianna University School of Medicine, Yokohama-City Seibu Hospital, Yokohama, Japan

Yuko Komase

398. Division of Hematology, Department of Internal Medicine, St Marianna University Yokohama-City Seibu Hospital, Yokohama, Japan

Fumiaki Sano

399. Division of Pulmonary Medicine, Department of Medicine, Tokai University School of Medicine, Tokai University School of Medicine, Tokyo, Japan

Koichiro Asano

400. Division of Pulmonary Medicine, Department of Medicine, Tokai University School of Medicine, Tokyo, Japan

Tsuyoshi Oguma

401. National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan

Masahiro Harada, Takeshi Takahashi & Takayuki Shibusawa

402. Department of Respiratory Medicine, Tokyo Medical University Hospital, Tokyo, Japan

Shinji Abe, Yuta Kono & Yuki Togashi

403. Department of Respiratory Medicine, Japanese Red Cross Medical Center, Tokyo, Japan

Takehiro Izumo, Minoru Inomata & Nobuyasu Awano

404. JA Toride Medical Hospital, Toride, Japan

Shinichi Ogawa, Tomouki Ogata & Shoichiro Ishihara

405. Japan Organization of Occupational Health and Safety Okayama Rosai Hospital, Okayama, Japan

Arihiko Kanehiro, Shinji Ozaki & Yasuko Fuchimoto

406. Emergency and Disaster Medicine, Graduate School of Medicine, Gifu University School of Medicine, Gifu, Japan

Yuichiro Kitagawa, Shozo Yoshida & Shinji Ogura

407. Niigata University, Niigata, Japan

Kei Nishiyama

408. National Hospital Organization Kyoto Medical Center, Kyoto, Japan

Kousuke Yoshida & Satoru Beppu

409. Research Institute for Diseases of the Chest, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

Satoru Fukuyama

410. Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan

Yoshihiro Eriguchi & Akiko Yonekawa

411. Department of Emergency and Critical Care Medicine, Tsukuba University, Tsukuba, Japan

Yoshiaki Inoue

412. Department of Nephrology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

Kunihiro Yamagata

413. Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

Shigeru Chiba

414. National Hospital Organization Tokyo Hospital, Tokyo, Japan

Osamu Narumoto, Hideaki Nagai & Nobuharu Ooshima

415. Fujioka General Hospital, Fujioka, Japan

Mitsuru Motegi

416. Division of Respiratory Medicine and Allergology, Department of Medicine, School of Medicine, Showa University, Tokyo, Japan

Hironori Sagara, Akihiko Tanaka & Shin Ohta

417. Department of Pulmonary Medicine, Fukushima Medical University, Fukushima, Japan

Yoko Shibata, Yoshinori Tanino & Yuki Sato

418. Kansai Electric Power Hospital, Osaka, Japan

Yuichiro Yamada, Takuya Hashino & Masato Shinoki

419. Kumamoto City Hospital, Kumamoto, Japan

Hajime Iwagoe

420. Department of Emergency and Critical Care Medicine, Tokyo Metropolitan Police Hospital, Tokyo, Japan

Tomonori Imamura

421. Department of Respiratory Medicine, International University of Health and Welfare, Shioya Hospital, Narita, Japan

Akira Umeda & Hisato Shimada

422. Department of Clinical Laboratory, International University of Health and Welfare, Shioya Hospital, Narita, Japan

Mayu Endo

423. National Hospital Organization Saitama Hospital, Saitama, Japan

Shinichi Hayashi, Mai Takahashi & Shigefumi Nakano

424. Department of Respiratory Medicine, Gunma University Graduate School of Medicine, Maebashi, Japan

Masakiyo Yatomi & Toshitaka Maeno

425. Department of Orthopedic Surgery, Tokyo Medical University, Ibaraki Medical Center, Tokyo, Japan

Tomoo Ishii

426. Department of Internal Medicine, Kiryu Kosei General Hospital, Kiryu, Japan

Mitsuyoshi Utsugi & Akihiro Ono

427. Daini Osaka Police Hospital, Osaka, Japan

Kensuke Kanaoka, Shoichi Ihara & Kiyoshi Komuta

428. Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands

Marike Boezen & Judith. M. Vonk

429. Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands

Anil P. S. Ori

430. Department of Genetics, University Medical Center Groningen,  
Groningen, The Netherlands

Anil P. S. Ori

431. Centre for Heart Lung Innovation, University of British Columbia,  
Vancouver, British Columbia, Canada

Ma' en Obeidat, Ana I. Hernández Cordero & Don D. Sin

432. Division of Respiratory Medicine, Faculty of Medicine, University of  
British Columbia, Vancouver, British Columbia, Canada

Don D. Sin

433. Institut Universitaire de Cardiologie et de Pneumologie de Québec,  
Université Laval, Quebec, Quebec, Canada

Yohan Bossé & Philippe Joubert

434. Department of Genetics and Genomic Sciences, Icahn School of  
Medicine at Mount Sinai, New York, NY, USA

Ke Hao, Robert P. Sebra, Daniel M. Jordan, Kyle Gettler, Kumardeep  
Chaudhary, Irene Salib, Marissa Zyndorf, Eric E. Schadt, Judy H.  
Cho, Yuval Itan & Ron Do

435. University of Washington, Global Health, Seattle, WA, USA

David Nickle

436. Gossamer Bio, San Diego, CA, USA

David Nickle

437. Department of Pathology and Medical Biology, University Medical  
Centre Groningen, University of Groningen, Groningen, The  
Netherlands

Wim Timens

438. GRIAC Research Institute, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands

Wim Timens & Maarten van den Berge

439. Department of Pulmonary Diseases, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands

Maarten van den Berge

440. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Yen-Chen Anne Feng & Josep Mercader

441. Harvard Medical School, Cambridge, MA, USA

James B. Meigs

442. Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA

James B. Meigs

443. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

Scott T. Weiss

444. Brigham and Women's Hospital, Boston, MA, USA

Elizabeth W. Karlson & Ann E. Woolley

445. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Jordan W. Smoller

446. Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

Shawn N. Murphy

447. Division of General Internal Medicine, Massachusetts General Hospital and Department of Medicine, Boston, MA, USA

James B. Meigs

448. Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

Brooke Wolford

449. Mount Sinai Clinical Intelligence Center, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Steven Ascolillo, Ryan C. Thompson, Noam D. Beckmann & Girish N. Nadkarni

450. Sema4, a Mount Sinai venture, Stamford, CT, USA

Robert P. Sebra & Eric E. Schadt

451. Mount Sinai Clinical Intelligence Center, Charles Bronfman Institute for Personalized Medicine, New York, NY, USA

Daniel M. Jordan

452. Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Bogdan Pasaniuc & Paul Boutros

453. Icahn Institute of Data Science and Genomics Technology, Icahn School of Medicine, New York, NY, USA

Ryan C. Thompson

454. Mount Sinai Clinical Intelligence Center, Icahn School of Medicine, New York, NY, USA

Ryan C. Thompson & Alexander W. Charney

455. Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Clive Hoggart, Sam Choi, Paul O'Reilly & Laura M. Huckins

456. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Kumardeep Chaudhary, Michael Preuss, Judy H. Cho, Yuval Itan, Ruth J. F. Loos, Girish N. Nadkarni & Ron Do

457. Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Gillian M. Belbin, Noura S. Abul-Husn & Eimear E. Kenny

458. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Michael Preuss & Ruth J. F. Loos

459. Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Clive Hoggart, Sam Choi, Paul O'Reilly & Laura M. Huckins

460. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Clive Hoggart, Sam Choi, Paul O'Reilly & Laura M. Huckins

461. Icahn School of Medicine at Mount Sinai, New York, NY, USA

Sam Choi & Paul O'Reilly

462. Department of Psychiatry, Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Alexander W. Charney

463. Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Ruth J. F. Loos

464. Department of Human Genetics, Center for Autism Research and Treatment, Institute for Precision Health, University of California Los Angeles, Los Angeles, CA, USA

Daniel H. Geschwind

465. The Hasso Plattner Institute of Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Girish N. Nadkarni

466. BioMe Phenomics Center, >Icahn School of Medicine at Mount Sinai, New York, NY, USA

Girish N. Nadkarni

467. Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Girish N. Nadkarni

468. Regeneron Genetics Center, Tarrytown, NY, USA

Manuel A. R. Ferreira, Goncalo R. Abecasis, Michael N. Cantor, Jack A. Kosmicki, Julie E. Horowitz, Aris Baras & Ashish Yadav

469. Phenomic Analytics & Clinical Data Core, Geisinger Health System, Danville, PA, USA

Joseph B. Leader & Matthew C. Gass

470. Department of Population Health Sciences, Geisinger Health System, Danville, PA, USA

Anne E. Justice, Geetha Chittoor & Navya Shilpa Josyula

471. Department of Molecular and Functional Genomics, Geisinger Health System, Danville, PA, USA

Dave J. Carey & Tooraj Mirshahi

472. Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Anurag Verma, Marylyn D. Ritchie, Daniel Rader, Shefali S. Verma, Anastasia Lucas & Yuki Bradford

473. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

Binglan Li

474. Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, USA

Cynthia M. Bulik

475. Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, USA

Cynthia M. Bulik

476. Institute of Neuroscience and Physiology, University of Gothenburg,  
Gothenburg, Sweden

Mikael Landén

477. Department of Medical Sciences, University of Turin, Turin, Italy

Alfredo Brusco

478. Department of Clinical and Biological Sciences, University of Turin,  
Orbassano, Italy

Giovanni Battista Ferrero

479. Department of Pediatrics, Department of Microbiology, Immunology  
and Molecular Genetics, University of California Los Angeles, Los  
Angeles, CA, USA

Manish J. Butte

480. University of Genova, Genova, Italy

Marcello Scala & Antonella Riva

481. Hopital Mont-Godinne, Yvoir, Belgium

Jean-François Rahier

482. Department of Molecular Medicine, University of Pavia, Pavia, Italy

Elisa Giorgio

483. Department of Public Health and Pediatric Sciences, University of  
Turin, Turin, Italy

Diana Carli

484. Qatar Biobank for Medical Research, Qatar Foundation Research,  
Development and Innovation, Qatar Foundation, Doha, Qatar

Nahla Afifi

485. Latvian Biomedical Research and Study Centre, Riga, Latvia

Janis Klovins, Vita Rovite, Raimonds Rescenko, Raitis  
Peculis & Monta Ustinova

486. Department of Neuroscience, Karolinska Institutet, Stockholm,  
Sweden

Hugo Zeberg

487. Max Planck Institute for Evolutionary Anthropology, Leipzig,  
Germany

Hugo Zeberg

488. Anaesthesiology and Intensive Care Medicine, Department of Surgical  
Sciences, Uppsala University, Uppsala, Sweden

Robert Frithiof, Michael Hultström & Miklos Lipcsey

489. Integrative Physiology, Department of Medical Cell Biology, Uppsala  
University, Uppsala, Sweden

Michael Hultström

490. Hedenstierna Laboratory, CIRRUS, Anaesthesiology and Intensive  
Care Medicine, Department of Surgical Sciences, Uppsala University,  
Uppsala, Sweden

Miklos Lipcsey

491. Department of Computer Science, School of Engineering, University  
of California Los Angeles, Los Angeles, CA, USA

Ruth Johnson

492. University of California Los Angeles, Los Angeles, CA, USA

Daniel H. Geschwind & Manish J. Butte

493. Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA, USA

Nelson Freimer

494. Division of Immunology, Allergy, and Rheumatology, University of California Los Angeles, Los Angeles, CA, USA

Manish J. Butte

495. Department of Psychiatry, University of California Los Angeles, Los Angeles, CA, USA

Daniel H. Geschwind

496. Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA

Daniel H. Geschwind

497. Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA, USA

Bogdan Pasaniuc

498. Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, USA

Bogdan Pasaniuc

499. Bioinformatics IDP, UCLA, Los Angeles, CA, USA

Yi Ding & Alec Chiu

500. Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Timothy S. Chang

501. Department of Urology, David Geffen School of Medicine at UCLA,  
Los Angeles, CA, USA

Paul Boutros

502. Queen Mary University, London, UK

Richard H. Scott

503. UCL Great Ormond Street Institute of Child Health, London, UK

Richard H. Scott

504. University of Cambridge, Cambridge, UK

Augusto Rendon

505. Big Data Institute, Nuffield Department of Population Health, Li Ka  
Shing Centre for Health Information and Discovery, University of  
Oxford, Oxford, UK

Daniel J. Wilson, Sarah G. Earle, Shang-Kuan Lin, Nicolas  
Arning, Jacob Armstrong & Justine K. Rudkin

506. Experimental Medicine Division, Nuffield Department of Medicine,  
John Radcliffe Hospital, University of Oxford, Oxford, UK

Derrick W. Crook & David H. Wyllie

507. Public Health England, Field Service, Addenbrooke's Hospital,  
Cambridge, UK

David H. Wyllie

508. Public Health England, Data and Analytical Services, National  
Infection Service, London, UK

Anne Marie O' Connell

509. Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA

Masahiro Kanai

510. Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA

Kristin Tsuo

511. Department of Clinical Research and Leadership, George Washington University, Washington, DC, USA

Shawneequa Callier

512. Department of Human Genetics, The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

Nicole Soranzo

513. Strangeways Research Laboratory, The National Institute for Health Research Blood and Transplant Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK

Nicole Soranzo

514. Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK

Nicole Soranzo

515. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Jing Hua Zhao, John Danesh, Emanuele Di Angelantonio & Adam S. Butterworth

516. British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK

John Danesh, Emanuele Di Angelantonio & Adam S. Butterworth

517. The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK

John Danesh, Emanuele Di Angelantonio & Adam S. Butterworth

518. Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

John Danesh, Emanuele Di Angelantonio & Adam S. Butterworth

519. Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK

John Danesh

520. Department of Epidemiology, Emory University Rollins School of Public Health, North Druid Hills, GA, USA

Yan V. Sun

521. Atlanta CA Health Care System, North Druid Hills, GA, USA

Yan V. Sun

522. Center for Population Genomics, MAVERIC, VA Boston Healthcare System, Boston, MA, USA

Jennifer E. Huffman, Christopher J. O'Donnell & Gina Peloso

523. MAVERIC, VA Boston Healthcare System, Boston, MA, USA

Kelly Cho, J. Michael Gaziano & Yuk-Lam Ho

524. Stanford University, Stanford, CA, USA

Phil Tsao

525. Palo Alto VA Healthcare System, Stanford, CA, USA

Phil Tsao

526. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

Gina Peloso

527. Department of Haematology, Central Hospital of Bolzano (SABES-ASDAA), Bolzano, Italy

Michael Mian

528. Laboratory of Clinical Pathology, Hospital of Bressanone (SABES-ASDAA), Bressanone, Italy

Federica Scaggiante

529. University of Alcalá, Centro de Investigación Biomédica en Red en Enfermedades Respiratorias (CIBERES), Madrid, Spain

Aurora Solier, David Jiménez & Rosa Nieto

530. Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Xiao Chang, Joseph R. Glessner & Hakon Hakonarson

531. Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Joseph R. Glessner & Hakon Hakonarson

532. Faculty of Medicine, University of Iceland, Reykjavik, Iceland

Hakon Hakonarson

533. Infectious Disease Unit, Hospital of Massa, Massa, Italy

Antonella Vincenti

534. Department of Clinical Medicine, Public Health, Life and Environment Sciences, University of L'Aquila, L'Aquila, Italy

Claudio Ferri & Davide Grassi

535. UOSD Laboratorio di Genetica Medica - ASL Viterbo, San Lorenzo, Italy

Gloria Pessina

536. Unit of Infectious Diseases, S. M. Annunziata Hospital, Florence, Italy

Massimo Di Pietro

537. Infectious Disease Unit, Hospital of Lucca, Lucca, Italy

Sauro Luchi, Chiara Barbieri, Donatella Acquilini, Elena Andreucci, Francesco Paciosi, Monica Poscente & Paola Petrocelli

538. Department of Clinical and Experimental Medicine, Infectious Diseases Unit, University of Pisa, Pisa, Italy

Giusy Tiseo & Marco Falcone

539. Clinic of Infectious Diseases, Catholic University of the Sacred Heart, Rome, Italy

Francesco Vladimiro Segala

540. Department of Diagnostic and Laboratory Medicine, Institute of Biochemistry and Clinical Biochemistry, Fondazione Policlinico

Universitario A. Gemelli IRCCS, Catholic University of the Sacred Heart, Rome, Italy

Silvia Baroni

541. Private University in the Principality of Liechtenstein, Triesen, Liechtenstein

Christoph Gassner

542. Digestive Diseases Unit, Virgen del Rocío University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain

Manuel Romero-Gómez

543. Department of Biochemistry, University Hospital Vall d'Hebron, Barcelona, Spain

Alba-Estela Garcia-Fernandez, Albert Blanco-Grau, Andrea Caballero-Garralda, Cristina Cea & Francisco Rodriguez-Frias

544. University of Sevilla, Sevilla, Spain

Enrique J. Calderón, Francisco J. Medrano, Javier Ampuero, Juan Delgado, Juan M. Guerrero, Rubén Morilla & Vicente Friaza

545. Instituto de Biomedicina de Sevilla, Sevilla, Spain

Douglas Maya-Miles, Carmen de la Horra, Enrique J. Calderón, Francisco J. Medrano, Javier Ampuero, Juan Delgado, Juan M. Guerrero, Rubén Morilla & Vicente Friaza

546. Hospital Universitario Virgen del Rocío de Sevilla, Sevilla, Spain

Douglas Maya-Miles, Carmen de la Horra, Enrique J. Calderón, Javier Ampuero, Juan Delgado, Juan M. Guerrero, Rocío Gallego-Durán, Rubén Morilla & Vicente Friaza

547. Consejo Superior de Investigaciones científicas, Madrid, Spain

Carmen de la Horra, Enrique J. Calderón, Francisco J. Medrano, Juan Delgado, Rubén Morilla & Vicente Friaza

548. Humanitas Clinical and Research Center, IRCCS, Milan, Italy

Claudio Angelini, Hayato Kurihara, Michele Ciccarelli, Monica Bocciolone, Paoletta Preatoni & Paolo Omodei

549. Immunohematology Department, Banc de Sang i Teixits, Autonomous University of Barcelona, Barcelona, Spain

Eduardo Muñiz-Díaz

550. August Pi i Sunyer Biomedical Research Institute, Hospital Clinic, University of Barcelona, Barcelona, Spain

Elena Sandoval, Fátima Aziz, Jose Ferrusquía-Acosta & Leticia Moreira

551. Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy

Francesco Blasi & Stefano Aliberti

552. Internal Medicine Department, Virgen del Rocío University Hospital, Sevilla, Spain

Francisco J. Medrano

553. Department of Biomedical Sciences, Humanitas University, Milan, Italy

Giulia Cardamone

554. Department Emergency, Anesthesia and Intensive Care, University Milano-Bicocca, Monza, Italy

Giuseppe Foti

555. Department of Medical Sciences, Università degli Studi di Torino,  
Turin, Italy

Giuseppe Matullo & Serena Aneli

556. Department of Medical Microbiology, Clinic of Laboratory Medicine,  
St Olav's Hospital, Trondheim, Norway

Jan Egil Afset

557. Department of Infectious Diseases, St Olav's Hospital, Trondheim  
University Hospital, Trondheim, Norway

Jan Kristian Damås

558. Department of Clinical and Molecular Medicine, NTNU, Trondheim,  
Norway

Jan Kristian Damås, Kari Risnes & Laura Rachele Bettini

559. Department of Research, St Olav's Hospital, Trondheim University  
Hospital, Trondheim, Norway

Kari Risnes & Laura Rachele Bettini

560. Institute of Parasitology and Biomedicine Lopez-Neyra, Granada,  
Spain

Javier Martín & Marialbert Acosta-Herrera

561. Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany

Jeanette Erdmann

562. German Research Center for Cardiovascular Research, partner site  
Hamburg-Lübeck-Kiel, Lübeck, Germany

Jeanette Erdmann

563. University Heart Center Lübeck, Lübeck, Germany

Jeanette Erdmann

564. Department of Research, Ostfold Hospital Trust, Gralum, Norway

Jonas Bergan

565. Pediatric Departement, Centro Tettamanti- European Reference Network (ERN) PaedCan, EuroBloodNet, MetabERN-University of Milano-Bicocca-Fondazione MBBM/Ospedale San Gerardo, Milan, Italy

Andrea Biondi & Mariella D' Angiò

566. Geminicenter for Sepsis Research, Institute of Circulation and Medical Imaging (ISB), NTNU, Trondheim, Norway

Erik Solligård & Lise Tuset Gustad

567. Clinic of Anesthesia and Intensive Care, St Olav's Hospital, Trondheim University Hospital, Trondheim, Norway

Erik Solligård

568. Clinic of Medicine and Rehabilitation, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

Lise Tuset Gustad

569. Stefan-Morsch-Stiftung, Birkenfeld, Germany

Marco Schaefer & Wolfgang Peter

570. Center of Bioinformatics, Biostatistics, and Bioimaging, School of Medicine and Surgery, University of Milano Bicocca, Milan, Italy

Maria G. Valsecchi

571. Phase 1 Research Centre, ASST Monza, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

Marina Cazzaniga

572. Pneumologia ASST-Monza, University of Milano-Bicocca, Milano, Italy

Paola Faverio

573. School of Medicine and Surgery, University of Milano-Bicocca, Milano, Italy

Paolo Bonfanti

574. Infectious Diseases Unit, San Gerardo Hospital, Monza, Italy

Paolo Bonfanti

575. SODIR-VHIR research group, Barcelona, Spain

Ricard Ferrer

576. Bioinformatics area, Fiundación progreso y Salud, Andalucia, Spain

Ximo Dopazo

# **Consortia**

## **COVID-19 Host Genetics Initiative**

- **COVID-19 Host Genetics InitiativeLeadership**

- Mari E. K. Niemi
- , Juha Karjalainen
- , Rachel G. Liao
- , Benjamin M. Neale
- , Mark Daly
- & Andrea Ganna

- **Writing group**

•

- **Writing group leaders**

- Mari E. K. Niemi
- , Gita A. Pathak
- , Shea J. Andrews
- & Masahiro Kanai

- **Writing group members**

- Kumar Veerapen
- , Israel Fernandez-Cadenas
- , Eva C. Schulte
- , Pasquale Striano
- , Minttu Marttila
- , Camelia Minica
- , Eirini Marouli
- , Mohd Anisul Karim
- , Frank R. Wendt
- , Jeanne Savage
- , Laura Sloofman
- , Guillaume Butler-Laporte
- , Han-Na Kim
- , Stavroula Kanoni
- , Yukinori Okada
- , Jinyoung Byun
- , Younghun Han
- , Mohammed Jashim Uddin
- , George Davey Smith
- , Cristen J. Willer
- & Joseph D. Buxbaum

- **Analysis group**

- **Manuscript analyses team leader**
  - Juha Karjalainen
- **Manuscript analyses team member: meta-analysis**
  - Juha Mehtonen
- **Manuscript analyses team member: heritability, methods and supplements**
  - Mari E. K. Niemi
- **Manuscript analyses team member: PHEWAS**
  - Gita A. Pathak
- **Manuscript analyses team member: Mendelian randomization**
  - Shea J. Andrews
- **Manuscript analyses team member: PC projection and gene prioritization**
  - Masahiro Kanai
- **Manuscript analyses team member: gene prioritization**
  - Hilary Finucane
- **Manuscript analyses team member: sensitivity analysis**
  - Mattia Cordioli

- 

- **Manuscript analyses team members: PC projection**

- Alicia R. Martin
    - & Wei Zhou

- **In silico analysis team members**

- Mattia Cordioli
    - , Bogdan Pasaniuc
    - , Hanna Julianne
    - , Hugues Aschard
    - , Huwenbo Shi
    - , Loic Yengo
    - , Renato Polimanti
    - , Maya Ghoussaini
    - , Jeremy Schwartzentuber
    - , Ian Dunham
    - & Joseph D. Buxbaum

- **Project management group**

- **Project management leader**

- Rachel G. Liao

- **Project management support**

- Karolina Chwialkowska
    - , Margherita Francescatto
    - , Amy Trankiem
    - & Mary K. Balaconis

- **Phenotype steering group**

- Lea Davis
- , Andrea Ganna
- , Sulggi Lee
- , James Priest
- , Alessandra Renieri
- , Vijay G. Sankaran
- , David van Heel
- , Patrick Deelen
- , J. Brent Richards
- , Tomoko Nakanishi
- , Les Biesecker
- , V. Eric Kerchberger
- & J. Kenneth Baillie

- **Data dictionary**

- Alessandra Renieri
- , Francesca Mari
- , Anna Bernasconi
- , Stefano Ceri Baillie
- & Arif Canakoglu

- **Scientific communication group**

- **Scientific communication leaders**

- Kumar Veerapen
    - & Brooke Wolford

- **Scientific communication members**

- Amy Trankiem
    - , Annika Faucon
    - , Atanu Kumar Dutta
    - , Claudia Schurmann
    - , Emi Harry

- , Ewan Birney
- , Gita A. Pathak
- , Huy Nguyen
- , Jamal Nasir
- , James Priest
- , Mari Kaunisto
- , Minttu Marttila
- , Matthew Solomonson
- , Nicole Dueker
- , Nirmal Vadgama
- , Shea J. Andrews
- , Sophie Limou
- , Rachel G. Liao
- & Andrea Ganna

- **Translators**

- Souad Rahmouni
- , Hamdi Mbarek
- , Dima Darwish
- , Atanu Kumar Dutta
- , Md Mesbah Uddin
- , Israel Fernandez-Cadenas
- , Raquel Albertos
- , Jordi Pérez-Tur
- , Ruolin Li
- , Lasse Folkersen
- , Ida Moltke
- , Nils Koelling
- , Eva C. Schulte
- , Claudia Schurmann
- , Alexander Teumer
- , Athanasios Kousathanas
- , Eirini Marouli
- , Alicia Utrilla
- , Ricardo A. Verdugo
- , Ruth Zárate

- , Carolina Medina-Gómez
- , David Gómez-Cabrero
- , Elena Carnero-Montoro
- , Jordi Pérez-Tur
- , Israel Fernandez-Cadenas
- , Carmen L. Cadilla
- , Andrés Moreno-Estrada
- , Adriana Garmendia
- , Leire Moya
- , Bahar Sedaghati-Khayat
- , Minttu Marttila
- , Palwendé Romuald Boua
- , Guillaume Butler-Laporte
- , Marie-Julie Favé
- , Laurent Francioli
- , Audrey Lemaçon
- , Sophie Limou
- , Isabelle Migeotte
- , Atanu Kumar Dutta
- , Sanjay Patel
- , Reka Varnai
- , Jozsef L. Szentpeteri
- , Csilla Sipeky
- , Francesca Colombo
- , Margherita Francescatto
- , Kathrin von Hohenstaufen
- , Pietro Lio
- , Costanza Vallerga
- , Pasquale Striano
- , Qingbo Wang
- , Masahiro Kanai
- , Yosuke Tanigawa
- , Tomoko Nakanishi
- , Hogune Im
- , Chulho Han
- , Han Song
- , Jiwoo Lim

- , Younhe Lee
- , Sugyeong Kim
- , Sangyoon Im
- , Biljana Atanasovska
- , Hajar Fauzan Ahmad
- , Kumar Veerapen
- , Cindy Boer
- , Philip Jansen
- , Lude Franke
- , Karolina Chwialkowska
- , Elżbieta Kaja
- , Dorota Pasko
- , Ingrid Kennis-Szilagyi
- , Sergey A. Kornilov
- , Vid Prijatelj
- , Ivana Prokić
- , Ilangkumaran Sivanadhan
- , Sarala Perumal
- , Sahar Esmaeeli
- , Nathaniel M. Pearson
- , Ruth Zárate
- & Mohd Anisul Karim

- **Website Development**

- **Website development leaders**

- Huy Nguyen
    - & Matthew Solomonson

- **23andMe**

- **Analysis team leader**

- Adam Auton

- **Data collection leaders**
  - 23andMe COVID-19 Team\*
  - , Janie F. Shelton
  - & Anjali J. Shastri
- **Analysis team members**
  - Teresa Filshtain-Sonmez
  - , Daniella Coker
  - , Antony Symons
  - , Jorge Esparza-Gordillo
  - , Stella Aslibekyan
  - & Jared O'Connell
- **Data collection members**
  - Chelsea Ye
  - & Catherine H. Weldon
- **23andMe COVID-19 Team**
  - Adam Auton
- **ACCOuNT**
  - **Analysis team leader**
    - Minoli Perera
  - **Data collection leaders**
    - Kevin O'Leary
    - , Matthew Tuck
    - , Travis O'Brien

- , David Meltzer
  - , Peter O'Donnell
  - & Edith Nutescu
- **Analysis team members**
    - Guang Yang
  - **Data collection members**
    - Cristina Alarcon
    - , Stefanie Herrmann
    - , Sophia Mazurek
    - , Jeff Banagan
    - , Zacharia Hamidi
    - , April Barbour
    - , Noora Raffat
    - & Diana Moreno
  - **Admin team member**
    - Paula Friedman
- **Amsterdam UMC COVID Study Group**
  - **Analysis team leader**
    - Bart Ferwerda
  - **Data collection leaders**
    - Diederik van de Beek
    - , Matthijs C. Brouwer
    - , Alexander P. J. Vlaar
    - & W. Joost Wiersinga

- **Analysis team members**

- Danielle Posthuma
- , Elleke Tissink
- , A. H. Koos Zwinderman
- & Emil Uffelmann

- **Data collection members**

- Michiel van Agtmael
- , Anne Geke Algera
- , Frank van Baarle
- , Diane Bax
- , Martijn Beudel
- , Harm Jan Bogaard
- , Marije Bomers
- , Peter I. Bonta
- , Lieuwe Bos
- , Michela Botta
- , Justin de Brabander
- , Godelieve de Bree
- , Sanne de Bruin
- , Marianna Bugiani
- , Esther Bulle
- , Osoul Chouchane
- , Alex Cloherty
- , Dave Dongelmans
- , Paul Elbers
- , Lucas Fleuren
- , Suzanne Geerlings
- , Bart Geerts
- , Theo Geijtenbeek
- , Armand Girbes
- , Bram Goorhuis
- , Martin P. Grobusch
- , Florianne Hafkamp
- , Laura Hagens

- , Jorg Hamann
- , Vanessa Harris
- , Robert Hemke
- , Sabine M. Hermans
- , Leo Heunks
- , Markus Hollmann
- , Janneke Horn
- , Joppe W. Hovius
- , Menno D. de Jong
- , Rutger Koning
- , Niels van Mourik
- , Jeannine Nellen
- , Esther J. Nossent
- , Frederique Paulus
- , Edgar Peters
- , Tom van der Poll
- , Bennedikt Preckel
- , Jan M. Prins
- , Jorinde Raasveld
- , Tom Reijnders
- , Michiel Schinkel
- , Marcus J. Schultz
- , Alex Schuurman
- , Kim Sigaloff
- , Marry Smit
- , Cornelis S. Stijnis
- , Willemke Stilma
- , Charlotte Teunissen
- , Patrick Thoral
- , Anissa Tsonas
- , Marc van der Valk
- , Denise Veelo
- , Heder de Vries
- , Michèle van Vugt
- , Dorien Wouters
- , René P. Minnaar
- , Adrie Kromhout

- , Kees W. J. van Uffelen
- & Ruud A. Wolterman

- **AncestryDNA COVID-19 Research Study**

- **Analysis team leader**
  - Genevieve Roberts
- **Data collection leader**
  - Danny Park
- **Admin team leader**
  - Catherine A. Ball
- **Analysis team members**
  - Marie Coignet
  - , Shannon McCurdy
  - , Spencer Knight
  - , Raghavendran Partha
  - & Brooke Rhead
- **Data collection members**
  - Miao Zhang
  - , Nathan Berkowitz
  - , Michael Gaddis
  - , Keith Noto
  - , Luong Ruiz
  - & Milos Pavlovic
- **Admin team members**

- Eurie L. Hong
- , Kristin Rand
- , Ahna Girshick
- , Harendra Guturu
- & Asher Haug Baltzell

- **BelCovid**

- **Analysis team leader**
  - Mari E. K. Niemi
- **Data collection leaders**
  - Isabelle Migeotte
  - , Souad Rahmouni
  - & Julien Guntz
- **Admin team leader**
  - Yves Beguin
- **Analysis team members**
  - Mattia Cordioli
  - , Sara Pigazzini
  - & Lindokuhle Nkambule
- **Data collection members**
  - Youssef Bouysran
  - , Adeline Busson
  - , Xavier Peyrassol
  - , Françoise Wilkin
  - , Bruno Pichon

- , Guillaume Smits
- , Isabelle Vandernoot
- , Jean-Christophe Goffard
- , Michel Georges
- , Michel Moutschen
- , Benoit Misset
- , Gilles Darcis
- , Julien Guiot
- , Laurent Jadot
- , Samira Azarzar
- , Patricia Dellot
- , Stéphanie Gofflot
- , Sabine Claassen
- , Axelle Bertrand
- , Gilles Parzibut
- , Mathilde Clarinval
- , Catherine Moermans
- , Olivier Malaise
- , Kamilia El Kandoussi
- , Raphaël Thonon
- , Pascale Huynen
- , Alyssia Mesdagh
- , Sofia Melo
- , Nicolas Jacques
- , Emmanuel Di Valentin
- , François Giroule
- , Alice Collignon
- , Coraline Radermecker
- , Marielle Lebrun
- , Alice Collignon
- , Hélène Perée
- , Samuel Latour
- , Olivia Barada
- , Judit Sanchez
- , Claire Josse
- , Bouchra Boujemla
- , Margot Meunier

- , Emeline Mariavelle
- , Sandy Anania
- , Hélène Gazon
- , Danusia Juszczak
- , Marjorie Fadeur
- , Séverine Camby
- , Christelle Meuris
- , Marie Thys
- , Jessica Jacques
- , Monique Henket
- , Philippe Léonard
- , Frederic Frippiat
- , Jean-Baptiste Giot
- , Anne-Sophie Sauvage
- , Christian Von Frenckell
- , Myriam Mni
- , Marie Wéry
- , Alicia Staderoli
- , Yasmine Belhaj
- & Bernard Lambermont

- **Biobanque Quebec COVID-19**

- **Analysis team leader**
- Tomoko Nakanishi

•

- **Data collection leader**
  - David R. Morrison
- **Admin team leaders**
  - Vincent Mooser
  - & J. Brent Richards
- **Analysis team members**
  - Guillaume Butler-Laporte
  - , Vincenzo Forgetta
  - & Rui Li
- **Data collection members**
  - Biswarup Ghosh
  - , Laetitia Laurent
  - , Alexandre Belisle
  - , Danielle Henry
  - , Tala Abdullah
  - , Olumide Adeleye
  - , Noor Mamlouk
  - , Nofar Kimchi
  - , Zaman Afrasiabi
  - , Nardin Rezk
  - , Branka Vulesevic
  - , Meriem Bouab
  - , Charlotte Guzman
  - , Louis Petitjean
  - , Chris Tselios
  - , Xiaoqing Xue

- , Jonathan Afilalo
- , Marc Afilalo
- , Maureen Oliveira
- , Bluma Brenner
- , Nathalie Brassard
- & Madeleine Durand

- **Admin team members**

- Erwin Schurr
- , Pierre Lepage
- , Jiannis Ragoussis
- , Daniel Auld
- , Michaël Chassé
- , Daniel E. Kaufmann
- , G. Mark Lathrop
- & Darin Adra

- **BioVU**

- **Analysis team leaders**

- Lea K. Davis
- , Nancy J. Cox
- & Jennifer E. Below

- **Analysis team members**

- Julia M. Sealock
- , Annika B. Faucon
- , Megan M. Shuey
- , Hannah G. Polikowsky
- , Lauren E. Petty
- , Douglas M. Shaw
- , Hung-Hsin Chen
- & Wanying Zhu

- **Bonn Study of COVID-19 Genetics**

- **Data collection leader**

- Kerstin U. Ludwig

- **Analysis team members**

- Julia Schröder
    - & Carlo Maj

- **Data collection members**

- Selina Rolker
    - , Markus M. Nöthen
    - , Julia Fazaal
    - , Verena Keitel
    - , Björn-Erik Ole Jensen
    - , Torsten Feldt
    - , Ingo Kurth
    - , Nikolaus Marx
    - , Michael Dreher
    - , Isabell Pink
    - , Markus Cornberg
    - , Thomas Illig
    - , Clara Lehmann
    - , Philipp Schommers
    - , Max Augustin
    - , Jan Rybniker
    - , Lisa Knopp
    - , Thomas Eggermann
    - , Sonja Volland
    - , Janine Altmüller
    - , Marc M. Berger
    - , Thorsten Brenner
    - , Anke Hinney

- , Oliver Witzke
- , Robert Bals
- , Christian Herr
- , Nicole Ludwig
- & Jörn Walter
- **CHRIS**
  - **Analysis team leader**
    - Christian Fuchsberger
  - **Data collection leaders**
    - Cristian Pattaro
    - & Alessandro De Grandi
  - **Admin team leader**
    - Peter Pramstaller
  - **Analysis team members**
    - David Emmert
    - , Roberto Melotti
    - & Luisa Foco
  - **Admin team members**
    - Deborah Mascalzoni
    - , Martin Gögele
    - , Francisco Domingues
    - & Andrew Hicks
- **Colorado Center for Personalized Medicine (CCPM)**

- **Analysis team leader**
  - Christopher R. Gignoux
- **Data collection leaders**
  - Stephen J. Wicks
  - & Kristy Crooks
- **Admin team leader**
  - Kathleen C. Barnes
- **Analysis team members**
  - Michelle Daya
  - , Jonathan Shortt
  - , Nicholas Rafaels
  - & Sameer Chavan
- **Columbia University COVID-19 Biobank**
  - **Analysis team leaders**
    - David B. Goldstein
    - & Krzysztof Kiryluk
  - **Data collection leaders**
    - Soumitra Sengupta
    - , Wendy Chung
    - & Muredach P. Reilly
  - **Analysis team members**

- Atlas Khan
  - , Chen Wang
  - , Gundula Povysil
  - , Nitin Bhardwaj
  - , Ali G. Gharavi
  - & Iuliana Ionita-Laza
- **Data collection members**
  - Ning Shang
  - , Sheila M. O'Byrne
  - , Renu Nandakumar
  - , Amritha Menon
  - , Yat S. So
  - & Eldad Hod
- **Admin team member**
  - Danielle Pendrick
- **Corea (Genetics of COVID-19-related Manifestation)**
  - **Analysis team leader**
    - Han-Na Kim
  - **Data collection leaders**
    - Soo-Kyung Park
    - , Hyung-Lae Kim
    - , Chang Kyung Kang
    - , Hyo-Jung Lee
    - & Kyoung-Ho Song
  - **Admin team leaders**

- Kyung Jae Yoon
  - & Nam-Jong Paik
- **Analysis team members**
    - Woojin Seok
    - & Heejun Yoon
  - **Data collection members**
    - Eun-Jeong Joo
    - , Yoosoo Chang
    - , Seungho Ryu
    - , Wan Beom Park
    - , Jeong Su Park
    - , Kyoung Un Park
    - , Sin Young Ham
    - , Jongtak Jung
    - , Eu Suk Kim
    - & Hong Bin Kim
- **COVID-19-Hostage**
    - **Analysis team leaders**
      - David Ellinghaus
      - , Frauke Degenhardt
      - , Mario Cáceres
      - , Simona Juzenas
      - & Tobias L. Lenz
    - **Data collection leaders**
      - Agustín Albillas
      - , Antonio Julià
      - , Bettina Heidecker

- , Eva C. Schulte
- , Federico Garcia
- , Florian Kurth
- , Florian Tran
- , Frank Hanses
- , Heinz Zoller
- , Jan C. Holter
- , Javier Fernández
- , Leif Erik Sander
- , Philip Rosenstiel
- , Philipp Koehler
- , Rafael de Cid
- , Rosanna Asselta
- , Stefan Schreiber
- , Ute Hehr
- , Daniele Prati
- , Guido Baselli
- , Luca Valenti
- , Luis Bujanda
- , Jesus M. Banales
- , Stefano Duga
- , Mauro D' Amato
- , Manuel Romero-Gómez
- , Maria Buti
- & Pietro Invernizzi

- **Admin team leaders**

- Andre Franke
- , Johannes R. Hov
- , Tom H. Karlsen
- , Trine Folseraas
- & Douglas Maya-Miles

- **Analysis team members**

- Ana Teles
- , Clinton Azuure
- , Eike Matthias Wacker
- , Florian Uellendahl-Werth
- , Hesham ElAbd
- , Jatin Arora
- , Jon Lerga-Jaso
- , Lars Wienbrandt
- , Malte Christoph Rühlemann
- , Mareike Wendorff
- , May Sissel Vadla
- , Ole Bernt Lenning
- , Onur Özer
- , Ronny Myhre
- , Soumya Raychaudhuri
- , Anja Tanck
- , Christoph Gassner
- , Georg Hemmrich-Stanisak
- , Jan Kässens
- , Maria E. Figuera Basso
- , Martin Schulzky
- , Michael Wittig
- , Nicole Braun
- , Tanja Wesse
- , Wolfgang Albrecht
- & Xiaoli Yi

- **Data collection members**

- Aaron Blandino Ortiz
- , Adolfo Garrido Chercoles
- , Agustín Ruiz
- , Alberto Mantovani
- , Aleksander Rygh Holten
- , Alena Mayer
- , Alessandro Cherubini
- , Alessandro Protti

- , Alessio Aghemo
- , Alessio Gerussi
- , Alfredo Ramirez
- , Alice Braun
- , Ana Barreira
- , Ana Lleo
- , Anders Benjamin Kildal
- , Andrea Ganna
- , Andreas Glück
- , Anna Carreras Nolla
- , Anna Latiano
- , Anne Ma Dyrhol-Riise
- , Antonio Muscatello
- , Antonio Voza
- , Ariadna Rando-Segura
- , Aurora Solier
- , Banasik Karina
- , Beatriz Cortes
- , Beatriz Mateos
- , Beatriz Nafria-Jimenez
- , Benedikt Schaefer
- , Carla Bellinghausen
- , Carlos Ferrando
- , Carmen Quereda
- , Carsten Skurk
- , Charlotte Thibeault
- , Christoph D. Spinner
- , Christoph Lange
- , Cinzia Hu
- , Claudio Cappadona
- , Cristiana Bianco
- , Cristina Sancho
- , Dag Arne Lihaug Hoff
- , Daniela Galimberti
- , David Jiménez
- , David Pestaña
- , David Toapanta

- , Elena Azzolini
- , Elio Scarpini
- , Elisa T. Helbig
- , Eloisa Urrechaga
- , Elvezia Maria Paraboschi
- , Emanuele Pontali
- , Enric Reverter
- , Enrique Navas
- , Eunate Arana
- , Félix García Sánchez
- , Ferruccio Ceriotti
- , Francesco Malvestiti
- , Francisco Mesonero
- , Gianni Pezzoli
- , Giuseppe Lamorte
- , Holger Neb
- , Ilaria My
- , Isabel Hernández
- , Itziar de Rojas
- , Iván Galván-Femenia
- , Jan Heyckendorf
- , Jan Rybníkář
- , Joan Ramon Badia
- , Jochen Schneider
- , Josune Goikoetxea
- , Julia Kraft
- , Karl Erik Müller
- , Karoline I. Gaede
- , Koldo García-Etxebarria
- , Kristian Tonby
- , Lars Heggelund
- , Laura Izquierdo-Sánchez
- , Lauro Sumoy
- , Lena J. Lippert
- , Leonardo Terranova
- , Lindokuhle Nkambule
- , Lucia Garbarino

- , Luis Téllez
- , Luisa Roade
- , Mahnoosh Ostadreza
- , Maider Intxausti
- , Manolis Kogevinas
- , Mari E. K. Niemi
- , María A. Gutiérrez-Stampa
- , Maria J. G. T. Vehreschild
- , Marta Marquié
- , Massimo Castoldi
- , Mattia Cordioli
- , Maurizio Cecconi
- , Mercè Boada
- , Michael J. Seilmaier
- , Michela Mazzocco
- , Miguel Rodríguez-Gandía
- , Natale Imaz Ayo
- , Natalia Blay
- , Nilda Martínez
- , Norwegian SARS-CoV-2 Study Group\*
- , Oliver A. Cornely
- , Orazio Palmieri
- , Paolo Tentorio
- , Pedro M. Rodrigues
- , Pedro P. España
- , Per Hoffmann
- , Petra Bacher
- , Phillip Suwalski
- , Raúl de Pablo
- , Rosa Nieto
- , Salvatore Badalamenti
- , Sandra Ciesek
- , Sara Bombace
- , Sara Pigazzini
- , Sibylle Wilfling
- , Søren Brunak
- , Stefanie Heilmann-Heimbach

- , The Humanitas COVID-19 Task Force\*
- , The Humanitas Gavazzeni COVID-19 Task Force\*
- , Stephan Ripke
- , Thomas Bahmer
- , Ulf Landmesser
- , Ulrike Protzer
- , Valeria Rimoldi
- , Vegard Skogen
- , Victor Andrade
- , Victor Moreno
- , Wolfgang Poller
- , Xavier Farre
- , Xiaomin Wang
- , Yascha Khodamoradi
- , Zehra Karadeniz
- , Adolfo de Salazar
- , Adriana Palom
- , Alba-Estela Garcia-Fernandez
- , Albert Blanco-Grau
- , Alberto Zanella
- , Alessandra Bandera
- , Almut Nebel
- , Andrea Biondi
- , Andrea Caballero-Garralda
- , Andrea Gori
- , Andreas Lind
- , Anna Ludovica Fracanzani
- , Anna Peschuck
- , Antonio Pesenti
- , Carmen de la Horra
- , Chiara Milani
- , Cinzia Paccapelo
- , Claudio Angelini
- , Cristina Cea
- , Eduardo Muñiz-Díaz
- , Elena Sandoval
- , Enrique J. Calderón

- , Erik Solligård
- , Fátima Aziz
- , Filippo Martinelli-Boneschi
- , Flora Peyvandi
- , Francesco Blasi
- , Francisco J. Medrano
- , Francisco Rodriguez-Frias
- , Fredrik Müller
- , Giacomo Grasselli
- , Giorgio Costantino
- , Giulia Cardamone
- , Giuseppe Foti
- , Giuseppe Matullo
- , Hayato Kurihara
- , Jan Egil Afset
- , Jan Kristian Damås
- , Javier Ampuero
- , Javier Martín
- , Jeanette Erdmann
- , Jonas Bergan
- , Siegfried Goerg
- , Jose Ferrusquía-Acosta
- , Jose Hernández Quero
- , Juan Delgado
- , Juan M. Guerrero
- , Kari Risnes
- , Laura Rachele Bettini
- , Leticia Moreira
- , Lise Tuset Gustad
- , Luigi Santoro
- , Luigia Scudeller
- , Mar Riveiro-Barciela
- , Marco Schaefer
- , Maria Carrabba
- , Maria G. Valsecchi
- , María Hernandez-Tejero
- , Marialbert Acosta-Herrera

- , Mariella D' Angiò
- , Marina Baldini
- , Marina Cazzaniga
- , Michele Ciccarelli
- , Monica Bocciolone
- , Monica Miozzo
- , Natalia Chueca
- , Nicola Montano
- , Paola Faverio
- , Paoletta Preatoni
- , Paolo Bonfanti
- , Paolo Omodei
- , Pedro Castro
- , Ricard Ferrer
- , Roberta Gualtierotti
- , Rocío Gallego-Durán
- , Rubén Morilla
- , Sammra Haider
- , Sara Marsal
- , Serena Aneli
- , Serena Pelusi
- , Silvano Bosari
- , Stefano Aliberti
- , Susanne Dudman
- , Tenghao Zheng
- , Tomas Pumarola
- , Trinidad Gonzalez Cejudo
- , Valter Monzani
- , Vicente Friaza
- , Wolfgang Peter
- & Ximo Dopazo

- **Norwegian SARS-CoV-2 Study Group**

- Tom H. Karlsen

- **Humanitas COVID-19 Task Force**

- Stefano Duga
- The Humanitas Gavazzeni COVID-19 Task Force
  - Stefano Duga
- Admin team members
  - Sandra May
  - & Marit M. Grimsrud
- deCODE
  - Analysis team leader
    - Daniel F. Gudbjartsson
  - Data collection leader
    - Kari Stefansson
  - Analysis team members
    - Patrick Sulem
    - , Gardar Sveinbjornsson
    - , Pall Melsted
    - , Gudmundur Norddahl
    - & Kristjan Helgi Swerford Moore
  - Data collection members
    - Unnur Thorsteinsdottir
    - & Hilma Holm

- **Determining the Molecular Pathways & Genetic Predisposition of the Acute Inflammatory Process Caused by SARS-CoV-2**
  - **Analysis team leader**
    - Marta E. Alarcón-Riquelme
  - **Data collection leader**
    - David Bernardo
  - **Analysis team member**
    - Manuel Martínez-Bueno
  - **Data collection member**
    - Silvia Rojo Rello
  - **Estonian Biobank**
    - **Analysis team leader**
      - Reedik Magi
    - **Data collection leader**
      - Lili Milani
    - **Admin team leader**
      - Andres Metspalu
  - **Analysis team members**

- Triin Laisk
- , Kristi Läll
- & Maarja Lepamets

- **Data collection members**

- Tõnu Esko
- , Ene Reimann
- , Paul Naaber
- , Edward Laane
- , Jaana Pesukova
- , Pärt Peterson
- , Kai Kisand
- , Jekaterina Tabri
- , Raili Allos
- , Kati Hensen
- , Joel Starkopf
- , Inge Ringmets
- , Anu Tamm
- & Anne Kallaste

- **Admin team members**

- Helene Alavere
- , Kristjan Metsalu
- & Mairo Puusepp

- **FinnGen**

- **Data collection members**

- Kati Kristiansson
- , Juha Karjalainen
- , Sami Koskelainen
- , Markus Perola
- , Kati Donner

- , Katja Kivinen
  - & Aarno Palotie
- **Admin team member**
  - Mari Kaunisto
- **FinnGen Admin team leader**
  - Aarno Palotie
- **Functional Host Genomics in Infectious Diseases (FHoGID)**
  - **Analysis team leader**
    - Carlo Rivolta
  - **Data collection leaders**
    - Pierre-Yves Bochud
    - , Stéphanie Bibert
    - , Noémie Boillat
    - , Semira Gonseth Nussle
    - & Werner Albrich
  - **Analysis team members**
    - Mathieu Quinodoz
    - & Dhryata Kamdar
  - **Data collection members**
    - Noémie Suh
    - , Dionysios Neofytos
    - , Véronique Erard

- , Cathy Voide
- , FHoGID\*
- , RegCOVID\*
- , P-PredictUs\*
- , SeroCOVID\*
- & CRiPSI\*

- **FHoGID**

- P. Y. Bochud
- , C. Rivolta
- , S. Bibert
- , M. Quinodoz
- , D. Kamdar
- , D. Neofytos
- , V. Erard
- , C. Voide
- , R. Friolet
- , P. Vollenweider
- , J. L. Pagani
- , M. Oddo
- , F. Meyer zu Bentrup
- , A. Conen
- , O. Clerc
- , O. Marchetti
- , A. Guillet
- , C. Guyat-Jacques
- , S. Foucras
- , M. Rime
- , J. Chassot
- , M. Jaquet
- , R. Merlet Viollet
- , Y. Lannepoudenx
- & L. Portopena

- **RegCOVID**

- P. Y. Bochud
- , P. Vollenweider
- , J. L. Pagani
- , F. Desgranges
- , P. Filippidis
- , B. Guéry
- , D. Haefliger
- , E. E. Kampouri
- , O. Manuel
- , A. Munting
- , M. Papadimitriou-Olivgeris
- , J. Regina
- , L. Rochat-Stettler
- , V. Suttels
- , E. Tadini
- , J. Tschopp
- , M. Van Singer
- & B. Viala

- **P-PredictUs**

- N. Boillat-Blanco
- , T. Brahier
- , O. Hügli
- , J. Y. Meuwly
- & O. Pantet

- **SérocoViD**

- S. Gonseth Nussle
- , M. Bochud
- , V. D'Acremont
- & S. Estoppey Younes

- **CRiPSI**

- W. C. Albrich
- , N. Suh
- , A. Cerny
- , L. O'Mahony
- , C. von Mering
- , P. Y. Bochud
- , M. Frischknecht
- , G.-R. Kleger
- , M. Filipovic
- , C. R. Kahlert
- , H. Wozniak
- , T. Rochat Negro
- , J. Pugin
- , K. Bouras
- , C. Knapp
- , T. Egger
- , A. Perret
- , P. Montillier
- , C. di Bartolomeo
- & B. Barda

- **GCAT Genomes For Life**

- **Analysis team leader**
  - Rafael de Cid
- **Data collection leaders**
  - Anna Carreras
  - , Victor Moreno
  - & Manolis Kogevinas
- **Analysis team members**
  - Iván Galván-Femenía

- , Natalia Blay
- , Xavier Farré
- & Lauro Sumoy

- **Data collection members**

- Beatriz Cortés
- , Josep Maria Mercader
- , Marta Guindo-Martinez
- , David Torrents
- , Judith Garcia-Aymerich
- , Gemma Castaño-Vinyals
- & Carlota Dobaño

- **GEN-COVID Multicenter Study**

- **Analysis team leaders**

- Marco Gori
- & Mari E. K. Niemi

- **Data collection leaders**

- Alessandra Renieri
- , Francesca Mari
- , Mario Umberto Mondelli
- , Francesco Castelli
- , Massimo Vaghi
- , Stefano Rusconi
- , Francesca Montagnani
- , Elena Bargagli
- , Federico Franchi
- , Maria Antonietta Mazzei
- , Luca Cantarini
- , Danilo Tacconi
- , Marco Feri

- , Raffaele Scala
- , Genni Spargi
- , Cesira Nencioni
- , Maria Bandini
- , Gian Piero Caldarelli
- , Maurizio Spagnesi
- , Anna Canaccini
- , Agostino Ognibene
- , Antonella D' Arminio Monforte
- , Massimo Girardis
- , Andrea Antinori
- , Daniela Francisci
- , Elisabetta Schiaroli
- , Pier Giorgio Scotton
- , Sandro Panese
- , Renzo Scaggiante
- , Matteo Della Monica
- , Mario Capasso
- , Giuseppe Fiorentino
- , Marco Castori
- , Filippo Aucella
- , Antonio Di Biagio
- , Luca Masucci
- , Serafina Valente
- , Marco Mandalà
- , Patrizia Zucchi
- , Ferdinando Giannattasio
- , Domenico A. Coviello
- , Cristina Mussini
- , Giancarlo Bosio
- , Luisa Tavecchia
- , Lia Crotti
- , Marco Rizzi
- , Maria Teresa La Rovere
- , Simona Sarzi-Braga
- , Maurizio Bussotti
- , Sabrina Ravaglia

- , Rosangela Artuso
- , Antonio Perrella
- , Davide Romani
- , Paola Bergomi
- , Emanuele Catena
- , Antonella Vincenti
- , Claudio Ferri
- , Davide Grassi
- , Gloria Pessina
- , Mario Tumbarello
- , Massimo Di Pietro
- , Ravaglia Sabrina
- , Sauro Luchi
- , Chiara Barbieri
- , Donatella Acquilini
- , Elena Andreucci
- , Francesco Paciosi
- , Francesco Vladimiro Segala
- , Giusy Tiseo
- , Marco Falcone
- , Mirjam Lista
- , Monica Poscente
- , Oreste De Vivo
- , Paola Petrocelli
- , Alessandra Guarnaccia
- , Silvia Baroni
- & Valentina Perticaroli

- **Admin team leaders**

- Simone Furini
- & Simona Dei

- **Analysis team members**

- Elisa Benetti

- , Nicola Picchiotti
- , Maurizio Sanarico
- , Stefano Ceri
- , Pietro Pinoli
- , Francesco Raimondi
- , Filippo Biscarini
- , Alessandra Stella
- , Mattia Bergomi
- , Kristina Zguro
- , Katia Capitani
- , Mattia Cordioli
- , Sara Pigazzini
- , Lindokuhle Nkambule
- & Marco Tanfoni

- **Data collection members**

- Chiara Fallerini
- , Sergio Daga
- , Margherita Baldassarri
- , Francesca Fava
- , Elisa Frullanti
- , Floriana Valentino
- , Gabriella Doddato
- , Annarita Giliberti
- , Rossella Tita
- , Sara Amitrano
- , Mirella Bruttini
- , Susanna Croci
- , Ilaria Meloni
- , Maria Antonietta Mencarelli
- , Caterina Lo Rizzo
- , Anna Maria Pinto
- , Giada Beligni
- , Andrea Tommasi
- , Laura Di Sarno
- , Maria Palmieri

- , Miriam Lucia Carriero
- , Diana Alaverdian
- , Nicola Iuso
- , Gabriele Inchingolo
- , Stefano Busani
- , Raffaele Bruno
- , Marco Vecchia
- , Mary Ann Belli
- , Stefania Mantovani
- , Serena Ludovisi
- , Eugenia Quiros-Roldan
- , Melania Degli Antoni
- , Isabella Zanella
- , Matteo Siano
- , Arianna Emiliozzi
- , Massimiliano Fabbiani
- , Barbara Rossetti
- , Giacomo Zanelli
- , Laura Bergantini
- , Miriana D'Alessandro
- , Paolo Cameli
- , David Bennet
- , Federico Anedda
- , Simona Marcantonio
- , Sabino Scolletta
- , Susanna Guerrini
- , Edoardo Conticini
- , Bruno Frediani
- , Chiara Spertilli
- , Alice Donati
- , Luca Guidelli
- , Marta Corridi
- , Leonardo Croci
- , Paolo Piacentini
- , Elena Desanctis
- , Silvia Cappelli
- , Agnese Verzuri

- , Valentina Anemoli
- , Alessandro Pancrazi
- , Maria Lorubbio
- , Esther Merlini
- , Federica Gaia Miraglia
- , Sophie Venturelli
- , Andrea Cossarizza
- , Alessandra Vergori
- , Arianna Gabrieli
- , Agostino Riva
- , Francesco Pachiosi
- , Francesca Andretta
- , Francesca Gatti
- , Saverio Giuseppe Parisi
- , Stefano Baratti
- , Carmelo Piscopo
- , Roberta Russo
- , Immacolata Andolfo
- , Achille Iolascon
- , Massimo Carella
- , Giuseppe Merla
- , Gabriella Maria Squeo
- , Pamela Raggi
- , Carmen Marciano
- , Rita Perna
- , Matteo Bassetti
- , Maurizio Sanguinetti
- , Alessia Giorli
- , Lorenzo Salerni
- , Pierpaolo Parravicini
- , Elisabetta Menatti
- , Tullio Trotta
- , Gabriella Coiro
- , Fabio Lena
- , Enrico Martinelli
- , Sandro Mancarella
- , Chiara Gabbi

- , Franco Maggiolo
- , Diego Ripamonti
- , Tiziana Bachetti
- , Claudia Suardi
- , Gianfranco Parati
- , Giordano Bottà
- , Paolo Di Domenico
- , Ilaria Rancan
- , Francesco Bianchi
- , Riccardo Colombo
- , Chiara Barbieri
- , Donatella Acquilini
- , Elena Andreucci
- , Francesco Pachiosi
- , Francesco Vladimiro Segala
- , Giusy Tiseo
- , Marco Falcone
- , Mirjam Lista
- , Monica Poscente
- , Oreste De Vivo
- , Paola Petrocelli
- , Alessandra Guarnaccia
- , Silvia Baroni
- & Valentina Perticaroli

- **Genes & Health**

- **Analysis team leader**
  - David A. van Heel
- **Data collection leader**
  - Karen A. Hunt
- **Admin team leader**

- Richard C. Trembath

•

- **Analysis team members**
  - Qin Qin Huang
  - & Hilary C. Martin
- **Data collection members**
  - Dan Mason
  - , Bhavi Trivedi
  - & John Wright
- **Admin team members**
  - Sarah Finer
  - , Genes & Health Research Team\*
  - & Christopher J. Griffiths
- **Genes & Health Research Team**
  - Shaheen Akhtar
  - , Mohammad Anwar
  - , Elena Arciero
  - , Samina Ashraf
  - , Gerome Breen
  - , Raymond Chung
  - , Charles J. Curtis
  - , Maharun Chowdhury
  - , Grainne Colligan
  - , Panos Deloukas
  - , Ceri Durham
  - , Sarah Finer
  - , Chris Griffiths
  - , Qin Qin Huang

- , Matt Hurles
- , Karen A. Hunt
- , Shapna Hussain
- , Kamrul Islam
- , Ahsan Khan
- , Amara Khan
- , Cath Lavery
- , Sang Hyuck Lee
- , Robin Lerner
- , Daniel MacArthur
- , Bev MacLaughlin
- , Hilary Martin
- , Dan Mason
- , Shefa Miah
- , Bill Newman
- , Nishat Safa
- , Farah Tahmasebi
- , Richard C. Trembath
- , Bhavi Trivedi
- , David A. van Heel
- & John Wright

- **Genes for Good**

- **Analysis team leader**
  - Albert V. Smith
- **Data collection members**
  - Andrew P. Boughton
  - , Kevin W. Li
  - , Jonathon LeFaive
  - & Aubrey Annis

- **Genetic determinants of COVID-19 complications in the Brazilian population**

- **Analysis team leader**

- Mari E. K. Niemi

- **Data collection leader**

- Cinthia E. Jannes

- **Admin team leaders**

- Jose E. Krieger
    - & Alexandre C. Pereira

- **Analysis team members**

- Mariliza Velho
    - , Emanuelle Marques
    - , Mattia Cordioli
    - , Sara Pigazzini
    - & Lindokuhle Nkambule

- **Data collection members**

- Isabella Ramos Lima
    - , Mauricio Teruo Tada
    - & Karina Valino

- **Genetic influences on severity of COVID-19 illness in Korea**

- **Analysis team leaders**

- Mark McCarthy

- & Carrie Rosenberger
- **Data collection leader**
  - Jong Eun Lee
- **Analysis team members**
  - Diana Chang
  - , Christian Hammer
  - , Julie Hunkapiller
  - , Anubha Mahajan
  - , Sarah Pendergrass
  - , Lara Sucheston-Campbell
  - & Brian Yaspan
- **Data collection members**
  - Hyun Soo Lee
  - , Eunsoon Shin
  - , Hye Yoon Jang
  - , Sunmie Kim
  - , Sungmin Kym
  - , Yeon-Sook Kim
  - , Hyeongseok Jeong
  - , Ki Tae Kwon
  - , Shin-Woo Kim
  - , Jin Yong Kim
  - , Young Rock Jang
  - , Hyun ah Kim
  - , Ji Yeon Lee
  - , Jeong Eun Lee
  - , Shinwon Lee
  - , Kang-Won Choe
  - , Yu Min Kang
  - , Sun Ha Jee

- & Keum Ji Jung
- **Genomic epidemiology of SARS-CoV-2 host genetics in coronavirus disease 2019**
  - **Data collection leaders**
    - Victoria Parikh
    - , Euan Ashley
    - , Matthew Wheeler
    - , Manuel Rivas
    - , Carlos Bustamante
    - , Benjamin Pinsky
    - , Phillip Febbo
    - , Kyle Farh
    - , Gary P. Schroth
    - & Francis deSouza
  - **Admin team leaders**
    - Karen Dalton
    - & Jeff Christle
  - **Analysis team members**
    - Christopher Deboever
    - , Sándor Szalma
    - , Yosuke Tanigawa
    - , Simone Rubinacci
    - & Olivier Delaneau
  - **Data collection members**
    - John Gorzynski
    - , Hannah de Jong
    - , Shirley Sutton

- , Nathan Youlton
- , Ruchi Joshi
- , David Jimenez-Morales
- , Christopher Hughes
- , David Amar
- , Alex Ioannidis
- , Steve Hershman
- , Anna Kirillova
- , Kinya Seo
- , Yong Huang
- , Massa Shoura
- , Nathan Hammond
- , Nathaniel Watson
- , Archana Raja
- , ChunHong Huang
- , Malaya Sahoo
- & Hannah Wang

- **Admin team member**
  - Jimmy Zhen
- **Genotek COVID-19 study**
  - **Analysis team leader**
    - Alexander Rakitko
  - **Admin team leader**
    - Valery Ilinsky
  - **Analysis team members**
    - Danat Yermakovich
    - , Iaroslav Popov

- , Alexander Chernitsov
- , Elena Kovalenko
- , Anna Krasnenko
- , Nikolay Plotnikov
- , Ivan Stetsenko
- & Anna Kim

- **Helix & Healthy Nevada Project Exome+ COVID-19 Phenotypes**

- **Analysis team leader**

- Elizabeth T. Cirulli

- **Analysis team members**

- Kelly M. Schiabor Barrett
    - , Alexandre Bolze
    - , Simon White
    - , Nicole L. Washington
    - & James T. Lu

- **Data collection members**

- Stephen Riffle
    - , Francisco Tanudjaja
    - , Xueqing Wang
    - , Jimmy M. Ramirez III
    - , Nicole Leonetti
    - , Efren Sandoval
    - , Iva Neveux
    - , Shaun Dabe
    - & Joseph J. Grzymski

- **24Genetics & IdiPaz Genomic Variants associated to COVID-19 infection outcome**

- **Analysis team leader**
    - Juan Ignacio Esteban Miñano
  - **Data collection leader**
    - Luis A. Aguirre
  - **Admin team leader**
    - Eduardo López-Collazo
  - **Analysis team members**
    - Manuel de la Mata Pazos
    - , Luciano Cerrato
    - & Lasse Folkersen
  - **Data collection members**
    - Roberto Lozano-Rodríguez
    - , José Avendaño-Ortiz
    - , Verónica Terrón Arcos
    - , Karla Marina Montalbán-Hernández
    - , Jaime Valentín Quiroga
    - & Alejandro Pascual-Iglesias
  - **Admin team members**
    - Charbel Maroun-Eid
    - & Alejandro Martín-Quirós
- **Japan Coronavirus Taskforce**
    - **Analysis team leaders**

- Ho Namkoong
- , Yukinori Okada
- & Seiya Imoto

- **Data collection leaders**

- Kazuhiko Katayama
- , Koichi Fukunaga
- , Yuko Kitagawa
- , Toshiro Sato
- , Naoki Hasegawa
- , Atsushi Kumanogoh
- , Akinori Kimura
- , Masumi Ai
- & Katsushi Tokunaga

- **Admin team leaders**

- Takanori Kanai
- , Satoru Miyano
- & Seishi Ogawa

- **Analysis team members**

- Ryuya Edahiro
- , Kyuto Sonehara
- , Yuya Shirai
- & Masahiro Kanai

- **Data collection members**

- Makoto Ishii
- , Hiroki Kabata
- , Katsunori Masaki
- , Hirofumi Kamata
- , Shinnosuke Ikemura

- , Shotaro Chubachi
- , Satoshi Okamori
- , Hideki Terai
- , Hiromu Tanaka
- , Atsuho Morita
- , Ho Lee
- , Takanori Asakura
- , Junichi Sasaki
- , Hiroshi Morisaki
- , Yoshifumi Uwamino
- , Kosaku Nanki
- , Yohei Mikami
- , Kazunori Tomono
- , Kazuto Kato
- , Fumihiko Matsuda
- , Meiko Takahashi
- , Nobuyuki Hizawa
- , Yoshito Takeda
- , Haruhiko Hirata
- , Takayuki Shiroyama
- , Satoru Miyawaki
- , Ken Suzuki
- , Yuichi Maeda
- , Takuro Nii
- , Yoshimi Noda
- , Takayuki Niitsu
- , Yuichi Adachi
- , Takatoshi Enomoto
- , Saori Amiya
- , Reina Hara
- , Kunihiko Takahashi
- , Tatsuhiko Anzai
- , Takanori Hasegawa
- , Satoshi Ito
- , Ryuji Koike
- , Akifumi Endo
- , Yuji Uchimura

- , Yasunari Miyazaki
- , Takayuki Honda
- , Tomoya Tateishi
- , Shuji Tohda
- , Naoya Ichimura
- , Kazunari Sonobe
- , Chihiro Sassa
- , Jun Nakajima
- , Yasuhito Nannya
- , Yosuke Omae
- , Kazuhisa Takahashi
- , Norihiro Harada
- , Makoto Hiki
- , Haruhi Takagi
- , Ai Nakamura
- , Etsuko Tagaya
- , Masatoshi Kawana
- , Ken Arimura
- , Takashi Ishiguro
- , Noboru Takayanagi
- , Taisuke Isono
- , Yotaro Takaku
- , Kenji Takano
- , Ryusuke Anan
- , Yukiko Nakajima
- , Yasushi Nakano
- , Kazumi Nishio
- , Soichiro Ueda
- , Reina Hayashi
- , Hiroki Tateno
- , Isano Hase
- , Shuichi Yoshida
- , Shoji Suzuki
- , Keiko Mitamura
- , Fumitake Saito
- , Tetsuya Ueda
- , Masanori Azuma

- , Tadao Nagasaki
- , Yoshinori Yasui
- , Yoshinori Hasegawa
- , Yoshikazu Mutoh
- , Takashi Yoshiyama
- , Tomohisa Shoko
- , Mitsuaki Kojima
- , Tomohiro Adachi
- , Motonao Ishikawa
- , Kenichiro Takahashi
- , Kazuyoshi Watanabe
- , Tadashi Manabe
- , Fumimaro Ito
- , Takahiro Fukui
- , Yohei Funatsu
- , Hidefumi Koh
- , Yoshihiro Hirai
- , Hidetoshi Kawashima
- , Atsuya Narita
- , Kazuki Niwa
- , Yoshiyuki Sekikawa
- , Fukuki Saito
- , Kazuhisa Yoshiya
- , Tomoyuki Yoshihara
- , Yusuke Suzuki
- , Sohei Nakayama
- , Keita Masuzawa
- , Koichi Nishi
- , Masaru Nishitsuji
- , Maiko Tani
- , Takashi Inoue
- , Toshiyuki Hirano
- , Keigo Kobayashi
- , Naoki Miyazawa
- , Yasuhiro Kimura
- , Reiko Sado
- , Takashi Ogura

- , Hideya Kitamura
- , Kota Murohashi
- , Ichiro Nakachi
- , Rie Baba
- , Daisuke Arai
- , Satoshi Fuke
- , Hiroshi Saito
- , Naota Kuwahara
- , Akiko Fujiwara
- , Takenori Okada
- , Tomoya Baba
- , Junya Noda
- , Shuko Mashimo
- , Kazuma Yagi
- , Tetsuya Shiomi
- , Mizuha Hashiguchi
- , Toshio Odani
- , Takao Mochimaru
- , Yoshitaka Oyamada
- , Nobuaki Mori
- , Namiki Izumi
- , Kaoru Nagata
- , Reiko Taki
- , Koji Murakami
- , Mitsuhiro Yamada
- , Hisatoshi Sugiura
- , Kentaro Hayashi
- , Tetsuo Shimizu
- , Yasuhiro Gon
- , Shigeki Fujitani
- , Tomoya Tsuchida
- , Toru Yoshida
- , Takashi Kagaya
- , Toshiyuki Kita
- , Satoru Sakagami
- , Yoshifumi Kimizuka
- , Akihiko Kawana

- , Yoshihiko Nakamura
- , Hiroyasu Ishikura
- , Tohru Takata
- , Takahide Kikuchi
- , Daisuke Taniyama
- , Morio Nakamura
- , Nobuhiro Kodama
- , Yasunari Kaneyama
- , Shunsuke Maeda
- , Yoji Nagasaki
- , Masaki Okamoto
- , Sayoko Ishihara
- , Akihiro Ito
- , Yusuke Chihara
- , Mayumi Takeuchi
- , Keisuke Onoi
- , Naozumi Hashimoto
- , Keiko Wakahara
- , Akira Ando
- , Makoto Masuda
- , Aya Wakabayashi
- , Hiroki Watanabe
- , Hisako Sageshima
- , Taka-Aki Nakada
- , Ryuzo Abe
- , Tadanaga Shimada
- , Kodai Kawamura
- , Kazuya Ichikado
- , Kenta Nishiyama
- , Masaki Yamasaki
- , Satoru Hashimoto
- , Yu Kusaka
- , Takehiko Ohba
- , Susumu Isogai
- , Minoru Takada
- , Hidenori Kanda
- , Yuko Komase

- , Fumiaki Sano
- , Koichiro Asano
- , Tsuyoshi Oguma
- , Masahiro Harada
- , Takeshi Takahashi
- , Takayuki Shibusawa
- , Shinji Abe
- , Yuta Kono
- , Yuki Togashi
- , Takehiro Izumo
- , Minoru Inomata
- , Nobuyasu Awano
- , Shinichi Ogawa
- , Tomouki Ogata
- , Shoichiro Ishihara
- , ARIHICO Kanehiro
- , Shinji Ozaki
- , Yasuko Fuchimoto
- , Yuichiro Kitagawa
- , Shozo Yoshida
- , Shinji Ogura
- , Kei Nishiyama
- , Kousuke Yoshida
- , Satoru Beppu
- , Satoru Fukuyama
- , Yoshihiro Eriguchi
- , Akiko Yonekawa
- , Yoshiaki Inoue
- , Kunihiro Yamagata
- , Shigeru Chiba
- , Osamu Narumoto
- , Hideaki Nagai
- , Nobuharu Ooshima
- , Mitsuru Motegi
- , Hironori Sagara
- , Akihiko Tanaka
- , Shin Ohta

- , Yoko Shibata
- , Yoshinori Tanino
- , Yuki Sato
- , Yuichiro Yamada
- , Takuya Hashino
- , Masato Shinoki
- , Hajime Iwagoe
- , Tomonori Imamura
- , Akira Umeda
- , Hisato Shimada
- , Mayu Endo
- , Shinichi Hayashi
- , Mai Takahashi
- , Shigefumi Nakano
- , Masakiyo Yatomi
- , Toshitaka Maeno
- , Tomoo Ishii
- , Mitsuyoshi Utsugi
- , Akihiro Ono
- , Kensuke Kanaoka
- , Shoichi Ihara
- & Kiyoshi Komuta

- **Lifelines**

- **Analysis team leader**
  - Lude Franke
- **Data collection leader**
  - Marike Boezen
- **Analysis team members**
  - Patrick Deelen

- , Annique Claringbould
  - , Esteban Lopera
  - , Robert Warmerdam
  - , Judith. M. Vonk
  - & Irene van Blokland
- **Data collection members**
    - Pauline Lanting
    - & Anil P. S. Ori
- **Lung eQTL Consortium**
    - **Data collection members**
      - Ma'en Obeidat
      - , Ana I. Hernández Cordero
      - , Don D. Sin
      - , Yohan Bossé
      - , Philippe Joubert
      - , Ke Hao
      - , David Nickle
      - , Wim Timens
      - & Maarten van den Berge
- **Mass General Brigham-Host Vulnerability to COVID-19**
    - **Analysis team leaders**
      - Yen-Chen Anne Feng
      - & Josep Mercader
    - **Data collection leaders**
      - Scott T. Weiss
      - , Elizabeth W. Karlson

- , Jordan W. Smoller
  - , Shawn N. Murphy
  - , James B. Meigs
  - & Ann E. Woolley
- **Admin team leader**
  - Robert C. Green
- **Data collection member**
  - Emma F. Perez
- **Michigan Genomics Initiative (MGI)**
  - **Analysis team leader**
    - Brooke Wolford
  - **Admin team leader**
    - Sebastian Zöllner
  - **Analysis team members**
    - Jiongming Wang
    - & Andrew Beck
- **Mount Sinai Health System COVID-19 Genomics Initiative**
  - **Analysis team leader**
    - Laura G. Sloofman
  - **Data collection leaders**

- Steven Ascolillo
- , Robert P. Sebra
- , Brett L. Collins
- & Tess Levy

- **Admin team leaders**

- Joseph D. Buxbaum
- & Stuart C. Sealfon

- **Analysis team members**

- Shea J. Andrews
- , Daniel M. Jordan
- , Ryan C. Thompson
- , Kyle Gettler
- , Kumardeep Chaudhary
- , Gillian M. Belbin
- , Michael Preuss
- , Clive Hoggart
- , Sam Choi
- & Slayton J. Underwood

- **Data collection members**

- Irene Salib
- , Bari Britvan
- , Katherine Keller
- , Lara Tang
- , Michael Peruggia
- , Liam L. Hiester
- , Kristi Niblo
- , Alexandra Aksentijevich
- , Alexander Labkowsky
- , Avromie Karp
- , Menachem Zlatopolsky

- & Marissa Zyndorf
- **Admin team members**
  - Alexander W. Charney
  - , Noam D. Beckmann
  - , Eric E. Schadt
  - , Noura S. Abul-Husn
  - , Judy H. Cho
  - , Yuval Itan
  - , Eimear E. Kenny
  - , Ruth J. F. Loos
  - , Girish N. Nadkarni
  - , Ron Do
  - , Paul O'Reilly
  - & Laura M. Huckins
- **MyCode Health Initiative**
  - **Analysis team leaders**
    - Manuel A. R. Ferreira
    - & Goncalo R. Abecasis
  - **Data collection leaders**
    - Joseph B. Leader
    - & Michael N. Cantor
  - **Admin team leaders**
    - Anne E. Justice
    - & Dave J. Carey
  - **Analysis team members**

- Geetha Chittoor
  - , Navya Shilpa Josyula
  - , Jack A. Kosmicki
  - , Julie E. Horowitz
  - & Aris Baras
- **Data collection members**
  - Matthew C. Gass
  - & Ashish Yadav
- **Admin team member**
  - Tooraj Mirshahi
- **Netherlands Twin Register**
  - **Analysis team leader**
    - Jouke Jan Hottenga
  - **Data collection leader**
    - Meike Bartels
  - **Admin team leader**
    - Eco J. C. de Geus
  - **Analysis team member**
    - Michel G. Nivard
- **Penn Medicine Biobank**

- **Analysis team leaders**
  - Anurag Verma
  - & Marylyn D. Ritchie
- **Admin team leader**
  - Daniel Rader
- **Analysis team members**
  - Binglan Li
  - , Shefali S. Verma
  - , Anastasia Lucas
  - & Yuki Bradford
- **Population controls**
  - **Analysis team leader**
    - Federico Zara
  - **Analysis team members**
    - Vincenzo Salpietro
    - , Marcello Scala
    - , Michele Iacomino
    - , Paolo Scudieri
    - & Renata Bocciardi
  - **Data collection members**
    - Carlo Minetti
    - , Antonella Riva
    - , Maria Stella Vari

- , Myriam Mni
  - , Jean-François Rahier
  - , Elisa Giorgio
  - , Federico Zara
  - & Diana Carli
- **Data collection leaders**
    - Pasquale Striano
    - , Edouad Louis
    - , Michel Georges
    - , Souad Rahmouni
    - , Cynthia M. Bulik
    - , Mikael Landén
    - , Alfredo Brusco
    - & Giovanni Battista Ferrero
  - **Admin team leaders**
    - Francesca Madia
    - & Bengt Fundín
- **Qatar Genome Program**
  - **Analysis team leader**
    - Hamdi Mbarek
  - **Data collection leader**
    - Said I. Ismail
  - **Analysis team members**
    - Chadi Saad

- & Yaser Al-Sarraj
- **Data collection members**
  - Radja Messai Badji
  - , Wadha Al-Muftah
  - , Asma Al Thani
  - & Nahla Afifi
- **Study of the COVID-19 host genetics in the population of Latvia**
  - **Analysis team leader**
    - Janis Klovins
  - **Data collection leader**
    - Vita Rovite
  - **Analysis team members**
    - Raimonds Rescenko
    - & Raitis Peculis
  - **Data collection member**
    - Monta Ustinova
- **The genetic predisposition to severe COVID-19**
  - **Analysis team leader**
    - Mari E. K. Niemi
  - **Data collection leader**

- Hugo Zeberg
- **Analysis team members**
  - Mattia Cordioli
  - , Sara Pigazzini
  - & Lindokuhle Nkambule
- **Data collection members**
  - Robert Frithiof
  - , Michael Hultström
  - & Miklos Lipcsey
- **UCLA Precision Health COVID-19 Host Genomics Biobank**
  - **Analysis team leader**
    - Ruth Johnson
  - **Data collection leader**
    - UCLA Health ATLAS & Data Mart Working Group\*
  - **UCLA Health ATLAS & Data Mart Working Group**
    - Daniel H. Geschwind
  - **Admin team leaders**
    - Nelson Freimer
    - , Manish J. Butte
    - , Daniel H. Geschwind
    - & Bogdan Pasaniuc

- **Analysis team members**
  - Yi Ding
  - , Alec Chiu
  - , Timothy S. Chang
  - & Paul Boutros
- **UK 100,000 Genomes Project (Genomics England)**
  - **Analysis team leader**
    - Loukas Moutsianas
  - **Data collection leaders**
    - Mark J. Caulfield
    - & Richard H. Scott
  - **Analysis team members**
    - Athanasios Kousathanas
    - , Dorota Pasko
    - , Susan Walker
    - , Alex Stuckey
    - , Christopher A. Odhams
    - & Daniel Rhodes
  - **Data collection members**
    - Tom Fowler
    - , Augusto Rendon
    - , Georgia Chan
    - & Prabhu Arumugam
- **UK Biobank**

- **Analysis team leaders**
  - Tomoko Nakanishi
  - , Konrad J. Karczewski
  - , Alicia R. Martin
  - , Daniel J. Wilson
  - & Chris A. Spencer
- **Data collection leaders**
  - Derrick W. Crook
  - , David H. Wyllie
  - & Anne Marie O'Connell
- **Admin team leader**
  - J. Brent Richards
- **Analysis team members**
  - Guillaume Butler-Laporte
  - , Vincenzo Forgetta
  - , Elizabeth G. Atkinson
  - , Masahiro Kanai
  - , Kristin Tsuo
  - , Nikolas Baya
  - , Patrick Turley
  - , Rahul Gupta
  - , Raymond K. Walters
  - , Duncan S. Palmer
  - , Gopal Sarma
  - , Matthew Solomonson
  - , Nathan Cheng
  - , Wenhan Lu
  - , Claire Churchhouse
  - , Jacqueline I. Goldstein

- , Daniel King
- , Wei Zhou
- , Cotton Seed
- , Mark J. Daly
- , Benjamin M. Neale
- , Hilary Finucane
- , Sam Bryant
- , F. Kyle Satterstrom
- , Gavin Band
- , Sarah G. Earle
- , Shang-Kuan Lin
- , Nicolas Arning
- & Nils Koelling

- **Data collection members**

- Jacob Armstrong
- & Justine K. Rudkin

- **Admin team members**

- Shawneequa Callier
- , Sam Bryant
- & Caroline Cusick

- **UK Blood Donors Cohort**

- **Analysis team leaders**

- Nicole Soranzo
- & Jing Hua Zhao

- **Data collection leaders**

- John Danesh
- & Emanuele Di Angelantonio

- **Analysis team member**
  - Adam S. Butterworth
- **VA Million Veteran Program (MVP)**
  - **Analysis team leaders**
    - Yan V. Sun
    - & Jennifer E. Huffman
  - **Data collection leader**
    - Kelly Cho
  - **Admin team leaders**
    - Christopher J. O'Donnell
    - , Phil Tsao
    - & J. Michael Gaziano
  - **Analysis team member**
    - Gina Peloso
  - **Data collection member**
    - Yuk-Lam Ho
- **Val Gardena**
  - **Analysis team leader**
    - Christian Fuchsberger

- **Data collection leader**
  - Michael Mian
- **Data collection member**
  - Federica Scaggiante
- **Admin team members**
  - Cristian Pattaro
  - & Peter Pramstaller
- **CHOP\_CAG**
  - Xiao Chang
  - , Joseph R. Glessner
  - & Hakon Hakonarson
- **GenOMICC/ISARIC4C**
  - **Data collection leaders**
    - J. Kenneth Baillie
    - , Peter J. McGuigan
    - , Luke Stephen Prockter Moore
    - , Marcela Paola Vizcaychipi
    - , Kathryn Hall
    - , Andy Campbell
    - , Ailastair Nichol
    - , Geraldine Ward
    - , Valerie Joan Page
    - , Malcolm G. Semple
    - , Kayode Adeniji
    - , Daniel Agranoff
    - , Ken Agwuh

- , Dhiraj Ail
- , Erin L. Aldera
- , Ana Alegria
- , Brian Angus
- , Abdul Ashish
- , Dougal Atkinson
- , Shahedal Bari
- , Gavin Barlow
- , Stella Barnass
- , Nicholas Barrett
- , Christopher Bassford
- , Sneha Basude
- , David Baxter
- , Michael Beadsworth
- , Jolanta Bernatoniene
- , John Berridge
- , Nicola Best
- , Pieter Bothma
- , David Chadwick
- , Robin Brittain-Long
- , Naomi Bulteel
- , Tom Burden
- , Andrew Burtenshaw
- , Vikki Caruth
- , David Chadwick
- , Duncan Chamber
- , Nigel Chee
- , Jenny Child
- , Srikanth Chukkambotla
- , Tom Clark
- , Paul Collini
- , Catherine Cosgrove
- , Jason Cupitt
- , Maria-Teresa Cutino-Moguel
- , Paul Dark
- , Chris Dawson
- , Samir Dervisevic

- , Phil Donnison
- , Sam Douthwaite
- , Andrew Drummond
- , Ingrid DuRand
- , Ahilanadan Dushianthan
- , Tristan Dyer
- , Cariad Evans
- , Chi Eziefula
- , Christopher Fegan
- , Adam Finn
- , Duncan Fullerton
- , Sanjeev Garg
- , Atul Garg
- , Effrossyni Gkrania-Klotsas
- , Jo Godden
- , Arthur Goldsmith
- , Clive Graham
- , Elaine Hardy
- , Stuart Hartshorn
- , Daniel Harvey
- , Peter Havalda
- , Daniel B. Hawcutt
- , Maria Hobrok
- , Luke Hodgson
- , Anil Hormis
- , Michael Jacobs
- , Susan Jain
- , Paul Jennings
- , Agilan Kaliappan
- , Vidya Kasipandian
- , Stephen Kegg
- , Michael Kelsey
- , Jason Kendall
- , Caroline Kerrison
- , Ian Kerslake
- , Oliver Koch
- , Gouri Koduri

- , George Koshy
- , Shondipon Laha
- , Steven Laird
- , Susan Larkin
- , Tamas Leiner
- , Patrick Lillie
- , James Limb
- , Vanessa Linnett
- , Jeff Little
- , Mark Lyttle
- , Michael MacMahon
- , Emily MacNaughton
- , Ravish Mankregod
- , Huw Masson
- , Elijah Matovu
- , Katherine McCullough
- , Ruth McEwen
- , Manjula Meda
- , Gary H. Mills
- , Jane Minton
- , Karl Ward
- , Mariyam Mirfenderesky
- , Kavya Mohandas
- , Quen Mok
- , James Moon
- , Elinoor Moore
- , Patrick Morgan
- , Craig Morris
- , Katherine Mortimore
- , Samuel Moses
- , Mbiye Mpenge
- , Rohinton Mulla
- , Michael Murphy
- , Megan Nagel
- , Thapas Nagarajan
- , Mark Nelson
- , Matthew K. O’Shea

- , Igor Otahal
- , Marlies Ostermann
- , Mark Pais
- , Selva Panchatsharam
- , Danai Papakonstantinou
- , Hassan Paraiso
- , Brij Patel
- , Natalie Pattison
- , Justin Pepperell
- , Mark Peters
- , Mandeep Phull
- , Stefania Pintus
- , Jagtur Singh Pooni
- , Frank Post
- , David Price
- , Rachel Prout
- , Nikolas Rae
- , Henrik Reschreiter
- , Tim Reynolds
- , Neil Richardson
- , Mark Roberts
- , Devender Roberts
- , Alistair Rose
- , Guy Rousseau
- , Brendan Ryan
- , Taranprit Saluja
- , Aarti Shah
- , Prad Shanmuga
- , Anil Sharma
- , Anna Shawcross
- , Jeremy Sizer
- , Manu Shankar-Hari
- , Richard Smith
- , Catherine Snelson
- , Nick Spittle
- , Nikki Staines
- , Tom Stambach

- , Richard Stewart
- , Pradeep Subudhi
- , Tamas Szakmany
- , Kate Tatham
- , Jo Thomas
- , Chris Thompson
- , Robert Thompson
- , Ascanio Tridente
- , Darell Tupper-Carey
- , Mary Twagira
- , Andrew Ustianowski
- , Nick Vallotton
- , Lisa Vincent-Smith
- , Shico Visuvanathan
- , Alan Vuylsteke
- , Sam Waddy
- , Rachel Wake
- , Andrew Walden
- , Ingeborg Welters
- , Tony Whitehouse
- , Paul Whittaker
- , Ashley Whittington
- , Padmasayee Papineni
- , Meme Wijesinghe
- , Martin Williams
- , Lawrence Wilson
- , Sarah Cole
- , Stephen Winchester
- , Martin Wiselka
- , Adam Wolverson
- , Daniel G. Wooton
- , Andrew Workman
- , Bryan Yates
- & Peter Young

- **Analysis team members**

- J. Kenneth Baillie
- , Rupert Beale
- , Andrew D. Bretherick
- , Mark J. Caulfield
- , Sara Clohisey
- , Max Head Fourman
- , James Furniss
- , Elvina Gountouna
- , Graeme Grimes
- , Chris Haley
- , David Harrison
- , Caroline Hayward
- , Sean Keating
- , Lucija Klaric
- , Paul Klenerman
- , Athanasios Kousathanas
- , Andy Law
- , Alison M. Meynert
- , Jonathan Millar
- , Loukas Moutsianas
- , Erola Pairo-Castineira
- , Nicholas Parkinson
- , Dorota Pasko
- , Chris P. Ponting
- , David J. Porteous
- , Konrad Rawlik
- , Anne Richmond
- , Kathy Rowan
- , Clark D. Russell
- , Richard H. Scott
- , Xia Shen
- , Barbara Shih
- , Albert Tenesa
- , Veronique Vitart
- , Susan Walker
- , Bo Wang
- , James F. Wilson

- , Yang Wu
- , Jian Yang
- , Zhijian Yang
- , Marie Zechner
- , Ranran Zhai
- , Chengqing Zheng
- , Lisa Norman
- , Riinu Pius
- , Thomas M. Drake
- , Cameron J. Fairfield
- , Stephen R. Knight
- , Kenneth A. Mclean
- , Derek Murphy
- , Catherine A. Shaw
- , Jo Dalton
- , Michelle Girvan
- , Egle Saviciute
- , Stephanie Roberts
- , Janet Harrison
- , Laura Marsh
- , Marie Connor
- , Sophie Halpin
- , Clare Jackson
- , Carrol Gamble
- , Gary Leeming
- , Andrew Law
- , Murray Wham
- , Sara Clohisey
- , Ross Hendry
- & James Scott-Brown

- **Data collection members**

- Colin Begg
- , Sara Clohisey
- , Charles Hinds
- , Antonia Ying Wai Ho

- , Peter W. Horby
- , Julian Knight
- , Lowell Ling
- , David Maslove
- , Danny McAuley
- , Jonathan Millar
- , Hugh Montgomery
- , Alistair Nichol
- , Peter J. M. Openshaw
- , Chris P. Ponting
- , Kathy Rowan
- , Malcolm G. Semple
- , Manu Shankar-Hari
- , Charlotte Summers
- , Timothy Walsh
- , Lisa Armstrong
- , Hayley Bates
- , Emma Dooks
- , Fiona Farquhar
- , Brigid Hairsine
- , C. McParland
- , Sophie Packham
- , Zoe Alldis
- , Raine Astin-Chamberlain
- , Fatima Bibi
- , Jack Biddle
- , Sarah Blow
- , Matthew Bolton
- , Catherine Borra
- , Ruth Bowles
- , Maudrian Burton
- , Yasmin Choudhury
- , David Collier
- , Amber Cox
- , Amy Easthope
- , Patrizia Ebano
- , Stavros Fotiadis

- , Jana Gurasashvili
- , Rosslyn Halls
- , Pippa Hartridge
- , Delordson Kallon
- , Jamila Kassam
- , Ivone Lancoma-Malcolm
- , Maninderpal Matharu
- , Peter May
- , Oliver Mitchelmore
- , Tabitha Newman
- , Mital Patel
- , Jane Pheby
- , Irene Pinzuti
- , Zoe Prime
- , Oleksandra Prysyzhna
- , Julian Shiel
- , Melanie Taylor
- , Carey Tierney
- , Suzanne Wood
- , Anne Zak
- , Olivier Zongo
- , Miranda Forsey
- , Agilan Kaliappan
- , Anne Nicholson
- , Joanne Riches
- , Mark Vertue
- , Christopher Wasson
- , Stephanie Finn
- , Jackie Green
- , Erin Collins
- , Bernadette King
- , Lina Grauslyte
- , Musarat Hussain
- , Mandeep Phull
- , Tatiana Pogreban
- , Lace Rosaroso
- , Erika Salciute

- , George Franke
- , Joanna Wong
- , Aparna George
- , Louise Akeroyd
- , Shereen Bano
- , Matt Bromley
- , Lucy Gurr
- , Tom Lawton
- , James Morgan
- , Kirsten Sellick
- , Deborah Warren
- , Brian Wilkinson
- , Janet McGowan
- , Camilla Ledgard
- , Amelia Stacey
- , Kate Pye
- , Ruth Bellwood
- , Michael Bentley
- , Maria Hobrok
- , Ronda Loosley
- , Heather McGuinness
- , Helen Tench
- , Rebecca Wolf-Roberts
- , Sian Gibson
- , Amanda Lyle
- , Fiona McNeela
- , Jayachandran Radhakrishnan
- , Alistair Hughes
- , Asifa Ali
- , Megan Brady
- , Sam Dale
- , Annalisa Dance
- , Lisa Gledhill
- , Jill Greig
- , Kathryn Hanson
- , Kelly Holdroyd
- , Marie Home

- , Diane Kelly
- , Ross Kitson
- , Lear Matapure
- , Deborah Melia
- , Samantha Mellor
- , Tonicha Nortcliffe
- , Jez Pinnell
- , Matthew Robinson
- , Lisa Shaw
- , Ryan Shaw
- , Lesley Thomis
- , Alison Wilson
- , Tracy Wood
- , Lee-Ann Bayo
- , Ekta Merwaha
- , Tahira Ishaq
- , Sarah Hanley
- , David Antcliffe
- , Dorota Banach
- , Stephen Brett
- , Phoebe Coghlan
- , Ziortza Fernandez
- , Anthony Gordon
- , Roceld Rojo
- , Sonia Sousa Arias
- , Maie Templeton
- , Rajeev Jha
- , Vinodh Krishnamurthy
- , Lai Lim
- , Rehana Bi
- , Barney Scholefield
- , Lydia Ashton
- , Alison Williams
- , Claire Cheyne
- , Anne Saunderson
- , Angela Allan
- , Felicity Anderson

- , Callum Kaye
- , Jade Liew
- , Jasmine Medhora
- , Teresa Scott
- , Erin Trumper
- , Adriana Botello
- , Petra Polgarova
- , Katerina Stroud
- , Eoghan Meaney
- , Megan Jones
- , Anthony Ng
- , Shruti Agrawal
- , Nazima Pathan
- , Deborah White
- , Esther Daubney
- , Kay Elston
- , Robert Parker
- , Amie Reddy
- , Ian Turner-Bone
- , Laura Wilding
- , Peter Harding
- , Reni Jacob
- , Cathy Jones
- , Craig Denmade
- , Maria Croft
- , Ian White
- , Rajeev Jha
- , Vinodh Krishnamurthy
- , Li Lim
- , Denise Griffin
- , Nycola Muchenje
- , Mcdonald Mupudzi
- , Richard Partridge
- , Jo-Anna Conyngham
- , Rachel Thomas
- , Mary Wright
- , Maria Alvarez Corral

- , Victoria Bastion
- , Daphene Clarke
- , Beena David
- , Harriet Kent
- , Rachel Lorusso
- , Gamu Lubimbi
- , Sophie Murdoch
- , Melchizedek Penacerrada
- , Alastair Thomas
- , Jennifer Valentine
- , Ana Vochin
- , Retno Wulandari
- , Brice Djeugam
- , Joy Dawson
- , Sweyn Garrioch
- , Melanie Tolson
- , Jonathan Aldridge
- , Laura Gomes de Almeida Martins
- , Jaime Carungcong
- , Sarah Beavis
- , Katie Dale
- , Rachel Gascoyne
- , Joanne Hawes
- , Kelly Pritchard
- , Lesley Stevenson
- , Amanda Whileman
- , Anne Cowley
- , Judith Highgate
- , Rikki Crawley
- , Abigail Crew
- , Mishell Cunningham
- , Allison Daniels
- , Laura Harrison
- , Susan Hope
- , Ken Inwerekbu
- , Sian Jones
- , Nicola Lancaster

- , Jamie Matthews
- , Alice Nicholson
- , Gemma Wray
- , Leonie Benham
- , Zena Bradshaw
- , Joanna Brown
- , Melanie Caswell
- , Jason Cupitt
- , Sarah Melling
- , Stephen Preston
- , Nicola Slawson
- , Emma Stoddard
- , Scott Warden
- , Edward Combes
- , Teishel Joefield
- , Sonja Monnery
- , Valerie Beech
- , Sallyanne Trotman
- , Bridget Hopkins
- , James Scriven
- , Laura Thrasyvoulou
- , Heather Willis
- , Susan Anderson
- , Janine Birch
- , Emma Collins
- , Kate Hammerton
- , Ryan O'Leary
- , Caroline Abernathy
- , Louise Foster
- , Andrew Gratrix
- , Vicky Martinson
- , Priyai Parkinson
- , Elizabeth Stones
- , Llucia Carbal-Ortega
- , Ritoo Kapoor
- , David Loader
- , Karen Castle

- , Craig Brandwood
- , Lara Smith
- , Richard Clark
- , Katie Birchall
- , Laurel Kolakaluri
- , Deborah Baines
- , Anila Sukumaran
- , Isheunesu Mapfunde
- , Megan Meredith
- , Lucy Morris
- , Lucy Ryan
- , Amy Clark
- , Julia Sampson
- , Cecilia Peters
- , Martin Dent
- , Margaret Langley
- , Saima Ashraf
- , Shuying Wei
- , Angela Andrew
- , Manish Chablani
- , Amy Kirkby
- , Kimberley Netherton
- , Michelle Bates
- , Jo Dasgin
- , Jaspreet Gill
- , Annette Nilsson
- , James Scriven
- , Elena Apetri
- , Cathrine Basikolo
- , Bethan Blackledge
- , Laura Catlow
- , Bethan Charles
- , Paul Dark
- , Reece Doonan
- , Jade Harris
- , Alice Harvey
- , Daniel Horner

- , Karen Knowles
- , Stephanie Lee
- , Diane Lomas
- , Chloe Lyons
- , Tracy Marsden
- , Danielle McLaughlan
- , Liam McMorrow
- , Jessica Pendlebury
- , Jane Perez
- , Maria Poulaka
- , Nicola Proudfoot
- , Melanie Slaughter
- , Kathryn Slevin
- , Melanie Taylor
- , Vicky Thomas
- , Danielle Walker
- , Angiy Michael
- , Matthew Collis
- , Martyn Clark
- , Martinaoulding
- , Edward Jude
- , Jacqueline McCormick
- , Oliver Mercer
- , Darsh Potla
- , Hafiz Rehman
- , Heather Savill
- , Victoria Turner
- , Miriam Davey
- , David Golden
- , Rebecca Seaman
- , Jodie Hunt
- , Joy Dearden
- , Emma Dobson
- , Andrew Drummond
- , Michelle Mulcahy
- , Sheila Munt
- , Grainne O'Connor

- , Jennifer Philbin
- , Chloe Rishton
- , Redmond Tully
- , Sarah Winnard
- , Lenka Cagova
- , Adama Fofano
- , Lucie Garner
- , Helen Holcombe
- , Sue Mepham
- , Alice Michael Mitchell
- , Lucy Mwaura
- , K. Praman
- , Alain Vuylsteke
- , Julie Zamikula
- , Miriam Davey
- , David Golden
- , Rebecca Seaman
- , Georgia Bercades
- , David Brealey
- , Ingrid Hass
- , Niall MacCallum
- , Gladys Martir
- , Eamon Raith
- , Anna Reyes
- , Deborah Smyth
- , Abigail Taylor
- , Rachel Anne Hughes
- , Helen Thomas
- , Alun Rees
- , Michaela Duskova
- , Janet Phipps
- , Suzanne Brooks
- , Michelle Edwards
- , Peter Alexander
- , Schvearn Allen
- , Joanne Bradley-Potts
- , Craig Brantwood

- , Jasmine Egan
- , Timothy Felton
- , Grace Padden
- , Luke Ward
- , Stuart Moss
- , Susannah Glasgow
- , Kate Beesley
- , Sarah Board
- , Agnieszka Kubisz-Pudelko
- , Alison Lewis
- , Jess Perry
- , Lucy Pippard
- , Di Wood
- , Clare Buckley
- , Alison Brown
- , Jane Gregory
- , Susan O'Connell
- , Tim Smith
- , Zakaula Belagodu
- , Bridget Fuller
- , Anca Gherman
- , Olumide Olufuwa
- , Remi Paramsothy
- , Carmel Stuart
- , Naomi Oakley
- , Charlotte Kamundi
- , David Tyl
- , Katy Collins
- , Pedro Silva
- , June Taylor
- , Laura King
- , Charlotte Coates
- , Maria Crowley
- , Phillipa Wakefield
- , Jane Beadle
- , Laura Johnson
- , Janet Sargeant

- , Madeleine Anderson
- , Catherine Jardine
- , Dewi Williams
- , Victoria Parris
- , Sheena Quaid
- , Ekaterina Watson
- , Julie Melville
- , Jay Naisbitt
- , Rosane Joseph
- , Maria Lazo
- , Olivia Walton
- , Alan Neal
- , Michaela Hill
- , Thogulava Kannan
- , Laura Wild
- , Elizabeth Allan
- , Kate Darlington
- , Ffyon Davies
- , Jack Easton
- , Sumit Kumar
- , Richard Lean
- , Daniel Menzies
- , Richard Pugh
- , Xinyi Qiu
- , Llinos Davies
- , Hannah Williams
- , Jeremy Scanlon
- , Gwyneth Davies
- , Callum Mackay
- , Joanne Lewis
- , Stephanie Rees
- , Samantha Coetzee
- , Alistair Gales
- , Igor Otahal
- , Meena Raj
- , Craig Sell
- , Helen Langton

- , Rachel Prout
- , Malcolm Watterson
- , Catherine Novis
- , Gill Arbane
- , Aneta Bociek
- , Sara Campos
- , Neus Grau
- , Tim Owen Jones
- , Rosario Lim
- , Martina Marotti
- , Marlies Ostermann
- , Manu Shankar-Hari
- , Christopher Whitton
- , Anthony Barron
- , Ciara Collins
- , Sundeep Kaul
- , Heather Passmore
- , Claire Prendergast
- , Anna Reed
- , Paula Rogers
- , Rajvinder Shokkar
- , Meriel Woodruff
- , Hayley Middleton
- , Oliver Polgar
- , Claire Nolan
- , Vicky Thwaites
- , Kanta Mahay
- , Chunda Sri-Chandana
- , Joslan Scherewode
- , Lorraine Stephenson
- , Sarah Marsh
- , Hollie Bancroft
- , Mary Bellamy
- , Margaret Carmody
- , Jacqueline Daglish
- , Faye Moore
- , Joanne Rhodes

- , Mirriam Sangombe
- , Salma Kadiri
- , James Scriven
- , Amanda Ayers
- , Wendy Harrison
- , Julie North
- , Anna Cavazza
- , Maeve Cockrell
- , Eleanor Corcoran
- , Maria Depante
- , Clare Finney
- , Ellen Jerome
- , Mark McPhail
- , Monalisa Nayak
- , Harriet Noble
- , Kevin O'Reilly
- , Evita Pappa
- , Rohit Saha
- , Sian Saha
- , John Smith
- , Abigail Knighton
- , Mandy Gill
- , Paul Paul
- , Valli Ratnam
- , Sarah Shelton
- , Inez Wynter
- , David Baptista
- , Rebecca Crowe
- , Rita Fernandes
- , Rosaleen Herdman-Grant
- , Anna Joseph
- , Adam Loveridge
- , India McKenley
- , Eriko Morino
- , Andres Naranjo
- , Richard Simms
- , Kathryn Sollesta

- , Andrew Swain
- , Harish Venkatesh
- , Jacyntha Khera
- , Jonathan Fox
- , Russell Barber
- , Claire Hewitt
- , Annette Hilldrith
- , Karen Jackson-Lawrence
- , Sarah Shepardson
- , Maryanne Wills
- , Susan Butler
- , Silvia Tavares
- , Amy Cunningham
- , Julia Hindale
- , Sarwat Arif
- , Linsha George
- , Sophie Twiss
- , David Wright
- , Maureen Holland
- , Natalie Keenan
- , Marc Lyons
- , Helen Wassall
- , Chris Marsh
- , Mervin Mahenthran
- , Emma Carter
- , Thomas Kong
- , Oluronke Adanini
- , Nikhil Bhatia
- , Maines Msiska
- , Miranda Forsey
- , Agilan Kaliappan
- , Anne Nicholson
- , Joanne Riches
- , Mark Virtue
- , Louise Mew
- , Esther Mwaura
- , Richard Stewart

- , Felicity Williams
- , Lynn Wren
- , Sara-Beth Sutherland
- , Ceri Battle
- , Elaine Brinkworth
- , Rachel Harford
- , Carl Murphy
- , Luke Newey
- , Tabitha Rees
- , Marie Williams
- , Sophie Arnold
- , David Brealey
- , John Hardy
- , Henry Houlden
- , Eleanor Moncur
- , Eamon Raith
- , Ambreen Tariq
- , Arianna Tucci
- , Karen Convery
- , Deirdre Fottrell-Gould
- , Lisa Hudig
- , Jocelyn Keshet-Price
- , Georgina Randell
- , Katie Stammers
- , Marwa Abdelrazik
- , Dhanalakshmi Bakthavatsalam
- , Munzir Elhassan
- , Arunkumar Ganesan
- , Anne Haldeos
- , Jeronimo Moreno-Cuesta
- , Dharam Purohit
- , Rachel Vincent
- , Kugan Xavier
- , Kumar Rohit
- , Frater Alasdair
- , Malik Saleem
- , Carter David

- , Samuel Jenkins
- , Zoe Lamond
- , Alanna Wall
- , Bryan Yates
- , Jessica Reynolds
- , Helen Campbell
- , Maria Thompsom
- , Steve Dodds
- , Stacey Duffy
- , Deborah Butcher
- , Susie O'Sullivan
- , Nicola Butterworth-Cowin
- , Bethan Deacon
- , Meg Hibbert
- , Carla Pothecary
- , Dariusz Tetla
- , Christopher Woodford
- , Latha Durga
- , Gareth Kennard-Holden
- , Laura Ortiz-Ruiz de Gordoa
- , Emily Peasgood
- , Claire Phillips
- , Denise Skinner
- , Jane Gaylard
- , Dee Mullan
- , Julie Newman
- , Ellie Davies
- , Lisa Roche
- , Sonia Sathe
- , Lutece Brimfield
- , Zoe Daly
- , David Pogson
- , Steve Rose
- , Amy Collins
- , Waqas Khaliq
- , Estefania Treus Gude
- , Louise Allen

- , Eva Beranova
- , Nikki Crisp
- , Joanne Deery
- , Tracy Hazelton
- , Alicia Knight
- , Carly Price
- , Sorrell Tilbey
- , Salah Turki
- , Sharon Turney
- , Julian Giles
- , Simon Booth
- , Gillian Bell
- , Katy English
- , Amro Katary
- , Louise Wilcox
- , Rachael Campbell
- , Noreen Clarke
- , Jonathan Whiteside
- , Mairi Mascarenhas
- , Avril Donaldson
- , Joanna Matheson
- , Fiona Barrett
- , Marianne O'Hara
- , Laura O'Keefe
- , Clare Bradley
- , Dawn Collier
- , Anil Hormis
- , Rachel Walker
- , Victoria Maynard
- , Tahera Patel
- , Matthew Smith
- , Srikanth Chukkambotla
- , Aayesha Kazi
- , Janice Hartley
- , Joseph Dykes
- , Muhammad Hijazi
- , Sarah Keith

- , Meherunnisa Khan
- , Janet Ryan-Smith
- , Philippa Springle
- , Jacqueline Thomas
- , Nick Truman
- , Samuel Saad
- , Dabheoc Coleman
- , Christopher Fine
- , Roseanna Matt
- , Bethan Gay
- , Jack Dalziel
- , Syamlan Ali
- , Drew Goodchild
- , Rhiannan Harling
- , Ravi Bhatterjee
- , Wendy Goddard
- , Chloe Davison
- , Stephen Duberly
- , Jeanette Hargreaves
- , Rachel Bolton
- , Shondipon Laha
- , Mark Verlander
- , Alexandra Williams
- , Helen Blackman
- , Ben Creagh-Brown
- , Sinead Donlon
- , Natalia Michalak-Glinska
- , Sheila Mtuwa
- , Veronika Pристопан
- , Armorel Salberg
- , Eleanor Smith
- , Sarah Stone
- , Charles Piercy
- , Jerik Verula
- , Dorota Burda
- , Rugia Montaser
- , Lesley Harden

- , Irving Mayangao
- , Cheryl Marriott
- , Paul Bradley
- , Celia Harris
- , Joshua Cooper
- , Cheryl Finch
- , Sarah Liderth
- , Alison Quinn
- , Natalia Waddington
- , Katy Fidler
- , Emma Tagliavini
- , Kevin Donnelly
- , Lynn Abel
- , Michael Brett
- , Brian Digby
- , Lisa Gemmell
- , James Hornsby
- , Patrick MacGoey
- , Pauline O'Neil
- , Richard Price
- , Natalie Rodden
- , Kevin Rooney
- , Radha Sundaram
- , Nicola Thomson
- , Rebecca Flanagan
- , Gareth Hughes
- , Scott Latham
- , Emma McKenna
- , Jennifer Anderson
- , Robert Hull
- , Kat Rhead
- , Debbie Branney
- , Jordan Frankham
- , Sally Pitts
- , Nigel White
- , Daniele Cristiano
- , Natalie Dormand

- , Zohreh Farzad
- , Mahitha Gummadi
- , Kamal Liyanage
- , Brijesh V. Patel
- , Sara Salmi
- , Geraldine Sloane
- , Vicky Thwaites
- , Mathew Varghese
- , Anelise C. Zborowski
- , Sarah Bean
- , Karen Burt
- , Michael Spivey
- , Christine Eastgate-Jackson
- , Helder Filipe
- , Daniel Martin
- , Amitaa Maharajh
- , Sara Mingo Garcia
- , Mark De Neef
- , Bethan Deacon
- , Ceri Lynch
- , Carla Pothecary
- , Lisa Roche
- , Gwenllian Sera Howe
- , Jayaprakash Singh
- , Keri Turner
- , Hannah Ellis
- , Natalie Stroud
- , Shiney Cherian
- , Sean Cutler
- , Anne Emma Heron
- , Anna Roynon-Reed
- , Tamas Szakmany
- , Gemma Williams
- , Owen Richards
- , Yusuf Cheema
- , Norfaizan Ahmad
- , Joann Barker

- , Kris Bauchmuller
- , Sarah Bird
- , Kay Cawthon
- , Kate Harrington
- , Yvonne Jackson
- , Faith Kibutu
- , Becky Lenagh
- , Shamiso Masuko
- , Gary H. Mills
- , Ajay Raithatha
- , Matthew Wiles
- , Jayne Willson
- , Helen Newell
- , Alison Lye
- , Lorenza Nwafor
- , Claire Jarman
- , Sarah Rowland-Jones
- , David Foote
- , Joby Cole
- , Roger Thompson
- , James Watson
- , Lisa Hesseldon
- , Irene Macharia
- , Luke Chetam
- , Jacqui Smith
- , Amber Ford
- , Samantha Anderson
- , Kathryn Birchall
- , Kay Housley
- , Sara Walker
- , Leanne Milner
- , Helena Hanratty
- , Helen Trower
- , Patrick Phillips
- , Simon Oxspring
- , Ben Donne
- , Emily Bevan

- , Jane Martin
- , Dawn Trodd
- , Geoff Watson
- , Caroline Wrey Brown
- , Lara Bunni
- , Claire Jennings
- , Monica Latif
- , Rebecca Marshall
- , Gayathri Subramanian
- , Nageswar Bandla
- , Minnie GellamUCHO
- , Michelle Davies
- , Christopher Thompson
- , Laura Ortiz-Ruiz de Gordoa
- , Emily Peasgood
- , Claire Phillips
- , Denise Skinner
- , Jane Gaylard
- , Dee Mullan
- , Julie Newman
- , Phil Donnison
- , Fiona Trim
- , Beena Eapen
- , Cecilia Ahmed
- , Balvinder Baines
- , Sarah Clamp
- , Julie Colley
- , Risna Haq
- , Anne Hayes
- , Jonathan Hulme
- , Samia Hussain
- , Sibet Joseph
- , Rita Kumar
- , Zahira Maqsood
- , Manjit Purewal
- , Ben Chandler
- , Kerry Elliott

- , Janine Mallinson
- , Alison Turnbull
- , Kathy Dent
- , Elizabeth Horsley
- , Muhammad Nauman Akhtar
- , Sandra Pearson
- , Dorota Potoczna
- , Sue Spencer
- , Hayley Blakemore
- , Borislava Borislavova
- , Beverley Faulkner
- , Emma Gendall
- , Elizabeth Goff
- , Kati Hayes
- , Matt Thomas
- , Ruth Worner
- , Kerry Smith
- , Deanna Stephens
- , Carlos Castro Delgado
- , Deborah Dawson
- , Lijun Ding
- , Georgia Durrant
- , Obiageri Ezeobu
- , Sarah Farnell-Ward
- , Abiola Harrison
- , Rebecca Kanu
- , Susannah Leaver
- , Elena Maccacari
- , Soumendu Manna
- , Romina Pepermans Saluzzio
- , Joana Queiroz
- , Tinashe Samakomva
- , Christine Sicat
- , Joana Texeira
- , Edna Fernandes Da Gloria
- , Ana Lisboa
- , John Rawlins

- , Jisha Mathew
- , Ashley Kinch
- , William James Hurt
- , Nirav Shah
- , Victoria Clark
- , Maria Thanasi
- , Nikki Yun
- , Kamal Patel
- , Alison Brown
- , Vikki Crickmore
- , Gabor Debreceni
- , Joy Wilkins
- , Liz Nicol
- , Iona Burn
- , Geraldine Hambrook
- , Katarina Manso
- , Ruth Penn
- , Pradeep Shanmugasundaram
- , Julie Tebbutt
- , Danielle Thornton
- , Anthony Rostron
- , Alistair Roy
- , Lindsey Woods
- , Sarah Cornell
- , Fiona Wakinshaw
- , Kimberley Rogerson
- , Jordan Jarman
- , Peter Anderson
- , Katie Archer
- , Karen Austin
- , Caroline Davis
- , Alison Durie
- , Olivia Kelsall
- , Jessica Thrush
- , Charlie Vigurs
- , Laura Wild
- , Hannah-Louise Wood

- , Helen Tranter
- , Alison Harrison
- , Nicholas Cowley
- , Michael McAlindon
- , Andrew Burtenshaw
- , Stephen Digby
- , Emma Low
- , Aled Morgan
- , Naiara Cother
- , Tobias Rankin
- , Sarah Clayton
- , Alex McCurdy
- , Suzanne Allibone
- , Roman Mary-Genetu
- , Vidya Kasipandian
- , Amit Patel
- , Ainho Mac
- , Anthony Murphy
- , Parisa Mahjoob
- , Roonak Nazari
- , Lucy Worsley
- , Andrew Fagan
- , Inthakab Ali Mohamed Ali
- , Karen Beaumont
- , Mark Blunt
- , Zoe Coton
- , Hollie Curgenven
- , Mohamed Elsaadany
- , Kay Fernandes
- , Sameena Mohamed Ally
- , Harini Rangarajan
- , Varun Sarathy
- , Sivarupan Selvanayagam
- , Dave Vedage
- , Matthew White
- , Jaime Fernandez-Roman
- , David O. Hamilton

- , Emily Johnson
- , Brian Johnston
- , Maria Lopez Martinez
- , Suleman Mulla
- , David Shaw
- , Alicia A. C. Waite
- , Victoria Waugh
- , Ingeborg D. Welters
- , Karen Williams
- , Thomas Bemand
- , Ethel Black
- , Arnold Dela Rosa
- , Ryan Howle
- , Shaman Jhanji
- , Ravishankar Rao Baikady
- , Kate Colette Tatham
- , Benjamin Thomas
- , Matthew Halkes
- , Pauline Mercer
- , Lorraine Thornton
- , Joe West
- , Tracy Baird
- , Jim Ruddy
- , Rosie Reece-Anthony
- , Mark Birt
- , Amanda Cowton
- , Andrea Kay
- , Melanie Kent
- , Kathryn Potts
- , Ami Wilkinson
- , Suzanne Naylor
- , Ellen Brown
- , Michele Clark
- , Sarah Purvis
- , Jade Cole
- , Michelle Davies
- , Rhys Davies

- , Donna Duffin
- , Helen Hill
- , Ben Player
- , Emma Thomas
- , Angharad Williams
- , Claire Marie Beith
- , Karen Black
- , Suzanne Clements
- , Alan Morrison
- , Dominic Strachan
- , Margaret Taylor
- , Michelle Clarkson
- , Stuart D'Sylva
- , Kathryn Norman
- , Tina Coventry
- , Susan Fowler
- , Michael MacMahon
- , Amanda McGregor
- , Ailbhe Brady
- , Rebekah Chan
- , Jeff Little
- , Shane McIvor
- , Helena Prady
- , Helen Whittle
- , Bijoy Mathew
- , Melanie Clapham
- , Rosemary Harper
- , Una Poultney
- , Polly Rice
- , Tim Smith
- , Rachel Mutch
- , Yolanda Baird
- , Aaron Butler
- , Indra Chadbourn
- , Linda Folkes
- , Heather Fox
- , Amy Gardner

- , Raquel Gomez
- , Gillian Hobden
- , Luke Hodgson
- , Kirsten King
- , Michael Margarson
- , Tim Martindale
- , Emma Meadows
- , Dana Raynard
- , Yvette Thirlwall
- , David Helm
- , Jordi Margalef
- , Sandra Greer
- , Karen Shuker
- , Ascanio Tridente
- , Sara Smuts
- , Joseph Duffield
- , Oliver Smith
- , Lewis Mallon
- , Watkins Claire
- , Isobel Birkinshaw
- , Joseph Carter
- , Kate Howard
- , Joanne Ingham
- , Rosie Joy
- , Harriet Pearson
- , Samantha Roche
- , Zoe Scott
- , Ellen Knights
- , Alicia Price
- , Alice Thomas
- , Chris Thorpe
- , Azmerelda Abraheem
- , Peter Bamford
- , Kathryn Cawley
- , Charlie Dunmore
- , Maria Faulkner
- , Rumanah Girach

- , Helen Jeffrey
- , Rhianna Jones
- , Emily London
- , Imrun Nagra
- , Farah Nasir
- , Hannah Sainsbury
- , Clare Smedley
- , Reena Khade
- , Ashok Sundar
- , George Tsinaslanidis
- , Teresa Behan
- , Caroline Burnett
- , Jonathan Hatton
- , Elaine Heeney
- , Atideb Mitra
- , Maria Newton
- , Rachel Pollard
- , Rachael Stead
- , Jenny Birch
- , Laura Bough
- , Josie Goodsell
- , Rebecca Tutton
- , Patricia Williams
- , Sarah Williams
- , Barbara Winter-Goodwin
- , Anne Cowley
- , Judith Highgate
- , Fiona Auld
- , Joanne Donnachie
- , Ian Edmond
- , Lynn Prentice
- , Nikole Runciman
- , Dario Salutous
- , Lesley Symon
- , Anne Todd
- , Patricia Turner
- , Abigail Short

- , Laura Sweeney
- , Euan Murdoch
- , Dhaneesha Senaratne
- , Karen Burns
- , Andrew Higham
- , Taya Anderson
- , Dan Hawcutt
- , Laura O’Malley
- , Laura Rad
- , Naomi Rogers
- , Paula Saunderson
- , Kathryn Sian Allison
- , Deborah Afolabi
- , Jennifer Whitbread
- , Dawn Jones
- , Rachael Dore
- , Liana Lankester
- , Nikitas Nikitas
- , Colin Wells
- , Bethan Stowe
- , Kayleigh Spencer
- , Susanne Cathcart
- , Katharine Duffy
- , Alex Puxty
- , Kathryn Puxty
- , Lynne Turner
- , Jane Ireland
- , Gary Semple
- , Peter Barry
- , Paula Hilltout
- , Jayne Evitts
- , Amanda Tyler
- , Joanne Waldron
- , Val Irvine
- , Benjamin Shelley
- , Olugbenga Akinkugbe
- , Alasdair Bamford

- , Emily Beech
- , Holly Belfield
- , Michael Bell
- , Charlene Davies
- , Gareth A. L. Jones
- , Tara McHugh
- , Hamza Meghari
- , Lauran O'Neill
- , Mark J. Peters
- , Samiran Ray
- , Ana Luisa Tomas
- , Amy Easthope
- , Claire Gorman
- , Abhinav Gupta
- , Elizabeth Timlick
- , Rebecca Brady
- , Stephen Bonner
- , Keith Hugill
- , Jessica Jones
- , Steven Liggett
- , Archana Bashyal
- , Neil Davidson
- , Paula Hutton
- , Stuart McKechnie
- , Jean Wilson
- , Neil Flint
- , Patel Rekha
- , Dawn Hales
- , Carina Cruz
- , Natalie Pattison
- , Shameer Gopal
- , Nichola Harris
- , Victoria Lake
- , Stella Metherell
- , Elizabeth Radford
- , Ian Clement
- , Bijal Patel

- , A. Gulati
- , Carole Hays
- , K. Webster
- , Anne Hudson
- , Andrea Webster
- , Elaine Stephenson
- , Louise McCormack
- , Victoria Slater
- , Rachel Nixon
- , Helen Hanson
- , Maggie Fearby
- , Sinead Kelly
- , Victoria Bridgett
- , Philip Robinson
- , Christine Almaden-Boyle
- , Pauline Austin
- , Louise Cabrelli
- , Stephen Cole
- , Matt Casey
- , Susan Chapman
- , Stephen Cole
- , Clare Whyte
- , Adam Brayne
- , Emma Fisher
- , Jane Hunt
- , Peter Jackson
- , Duncan Kaye
- , Nicholas Love
- , Juliet Parkin
- , Victoria Tuckey
- , Lynne van Koutrik
- , Sasha Carter
- , Benedict Andrew
- , Louise Findlay
- , Katie Adams
- , Michelle Bruce
- , Karen Connolly

- , Tracy Duncan
- , Helen T.-Michael
- , Gabriella Lindergard
- , Samuel Hey
- , Claire Fox
- , Jordan Alfonso
- , Laura Jayne Durrans
- , Jacinta Guerin
- , Bethan Blackledge
- , Jade Harris
- , Martin Hruska
- , Ayaa Eltayeb
- , Thomas Lamb
- , Tracey Hodgkiss
- , Lisa Cooper
- , Joanne Rothwell
- , Catherine Dennis
- , Alastair McGregor
- , Victoria Parris
- , Sinduya Srikanan
- , Anisha Sukha
- , Kim Davies
- , Linda O'Brien
- , Zohra Omar
- , Igor Otahal
- , Emma Perkins
- , Tracy Lewis
- , Isobel Sutherland
- , Hollie Brooke
- , Sarah Buckley
- , Jose Cebrian Suarez
- , Ruth Charlesworth
- , Karen Hansson
- , John Norris
- , Alice Poole
- , Alastair Rose
- , Rajdeep Sandhu

- , Brendan Sloan
- , Elizabeth Smithson
- , Muthu Thirumaran
- , Veronica Wagstaff
- , Alexandra Metcalfe
- , Julie Camsooksai
- , Charlotte Humphrey
- , Sarah Jenkins
- , Henrik Reschreiter
- , Beverley Wadams
- , Yasmin DeAth
- , Colene Adams
- , Anita Agasou
- , Tracie Arden
- , Amy Bowes
- , Pauline Boyle
- , Mandy Beekes
- , Heather Button
- , Nigel Capps
- , Mandy Carnahan
- , Anne Carter
- , Danielle Childs
- , Denise Donaldson
- , Kelly Hard
- , Fran Hurford
- , Yasmin Hussain
- , Ayesha Javaid
- , James Jones
- , Sanal Jose
- , Michael Leigh
- , Terry Martin
- , Helen Millward
- , Nichola Motherwell
- , Rachel Rikunenko
- , Jo Stickley
- , Julie Summers
- , Louise Ting

- , Helen Tivenan
- , Louise Tonks
- , Rebecca Wilcox
- , Maria Bokhari
- , Vanessa Linnett
- , Rachael Lucas
- , Wendy McCormick
- , Jenny Ritzema
- , Amanda Sanderson
- , Helen Wild
- , Nicola Baxter
- , Steven Henderson
- , Sophie Kennedy-Hay
- , Christopher McParland
- , Laura Rooney
- , Malcolm Sim
- , Gordan McCreath
- , Mark Brunton
- , Jess Caterson
- , Holly Coles
- , Matthew Frise
- , Sabi Gurung Rai
- , Nicola Jacques
- , Liza Keating
- , Emma Tilney
- , Shauna Bartley
- , Parminder Bhui
- , Charlotte Downes
- , Kathleen Holding
- , Katie Riches
- , Mary Hilton
- , Mel Hayman
- , Deepak Subramanian
- , Priya Daniel
- , Letizia Zitter
- , Sarah Benyon
- , Suzie Marriott

- , Linda Park
- , Samantha Keenan
- , Elizabeth Gordon
- , Helen Quinn
- , Kizzy Baines
- , Gillian Andrew
- , J. Kenneth Baillie
- , Lucy Barclay
- , Marie Callaghan
- , Rachael Campbell
- , Sarah Clark
- , Dave Hope
- , Lucy Marshall
- , Corrienne McCulloch
- , Kate Briton
- , Jo Singleton
- , Sophie Birch
- , Andrew Higham
- , Kerry Simpson
- , Jayne Craig
- , Carrie Demetriou
- , Charlotte Eckbad
- , Sarah Hierons
- , Lucy Howie
- , Sarah Mitchard
- , Lidia Ramos
- , Alfredo Serrano-Ruiz
- , Katie White
- , Fiona Kelly
- , Vishal Amin
- , Elena Anastasescu
- , Vikram Anumakonda
- , Komala Karthik
- , Rizwana Kausar
- , Karen Reid
- , Jacqueline Smith
- , Janet Imeson-Wood

- , Arianna Bellini
- , Jade Bryant
- , Anton Mayer
- , Amy Pickard
- , Nicholas Roe
- , Jason Sowter
- , Alex Howlett
- , Kristine Criste
- , Rebecca Cusack
- , Kim Golder
- , Hannah Golding
- , Oliver Jones
- , Samantha Leggett
- , Michelle Male
- , Martyna Marani
- , Kirsty Prager
- , Toran Williams
- , Belinda Roberts
- , Karen Salmon
- , Prisca Gondo
- , B. Hadebe
- , Abdul Kayani
- , Bridgett Masunda
- , Ashar Ahmed
- , Anna Morris
- , Srinivas Jakkula
- , Kate Long
- , Simon Whiteley
- , Elizabeth Wilby
- , Bethan Ogg
- , Sam Moultrie
- , M. Odam
- , Jeremy Bewley
- , Zoe Garland
- , Lisa Grimmer
- , Bethany Gumbrill
- , Rebekah Johnson

- , Katie Sweet
- , Denise Webster
- , Georgia Efford
- , Sara Bennett
- , Emma Goodwin
- , Matthew Jackson
- , Alissa Kent
- , Clare Tibke
- , Wiesia Woodyatt
- , Ahmed Zaki
- , Amelia Daniel
- , Joanne Finn
- , Rajnish Saha
- , Nikki Staines
- , Amy Easthope
- , Pamela Bremmer
- , J. Allan
- , T. Geary
- , Gordon Houston
- , A. Meikle
- , P. O'Brien
- , Dina Bell
- , Rosalind Boyle
- , Katie Douglas
- , Lynn Glass
- , Emma Lee
- , Liz Lennon
- , Austin Rattray
- , Rob Charnock
- , Denise McFarland
- , Denise Cosgrove
- , Ben Attwood
- , Penny Parsons
- , Siobhain Carmody
- , Metod Oblak
- , Monica Popescu
- , Mini Thankachen

- , Rosie Baruah
- , Sheila Morris
- , Susie Ferguson
- , Amy Shepherd
- , Abdelhakim Altabaibeh
- , Ana Alvaro
- , Kayleigh Gilbert
- , Louise Ma
- , Loreta Mostoles
- , Chetan Parmar
- , Kathryn Simpson
- , Champa Jetha
- , Lauren Booker
- , Anezka Pratley
- , Tracey Cosier
- , Gemma Millen
- , Neil Richardson
- , Natasha Schumacher
- , Heather Weston
- , James Rand
- , Beatrice Alex
- , Benjamin Bach
- , Wendy S. Barclay
- , Debby Bogaert
- , Meera Chand
- , Graham S. Cooke
- , Annemarie B. Docherty
- , Jake Dunning
- , Ana da Silva Filipe
- , Tom Fletcher
- , Christoper A. Green
- , Ewen M. Harrison
- , Julian A. Hiscox
- , Samreen Ijaz
- , Saye Khoo
- , Paul Klenerman
- , Andrew Law

- , Wei Shen Lim
- , Alexander J. Mentzer
- , Laura Merson
- , Alison M. Meynert
- , Mahdad Noursadeghi
- , Shona C. Moore
- , Massimo Palmarini
- , William A. Paxton
- , Georgios Pollakis
- , Nicholas Price
- , Andrew Rambaut
- , David L. Robertson
- , Clark D. Russell
- , Vanessa Sancho-Shimizu
- , Janet T. Scott
- , Thushan de Silva
- , Louise Sigfrid
- , Tom Solomon
- , Shiranee Sriskandan
- , David Stuart
- , Richard S. Tedder
- , Emma C. Thomson
- , A. A. Roger Thompson
- , Ryan S. Thwaites
- , Lance C. W. Turtle
- , Rishi K. Gupta
- , Carlo Palmieri
- , Olivia V. Swann
- , Maria Zambon
- , Marc-Emmanuel Dumas
- , Julian L. Griffin
- , Zoltan Takats
- , Kanta Chechi
- , Petros Andrikopoulos
- , Anthonia Osagie
- , Michael Olanipekun
- , Sonia Liggi

- , Matthew R. Lewis
- , Gonçalo dos Santos Correia
- , Caroline J. Sands
- , Panteleimon Takis
- , Lynn Maslen
- , William Greenhalf
- , Victoria Shaw
- , Sarah E. McDonald
- , Seán Keating
- , Katie A. Ahmed
- , Jane A. Armstrong
- , Milton Ashworth
- , Innocent G. Asiimwe
- , Siddharth Bakshi
- , Samantha L. Barlow
- , Laura Booth
- , Benjamin Brennan
- , Katie Bullock
- , Benjamin W. A. Catterall
- , Jordan J. Clark
- , Emily A. Clarke
- , Louise Cooper
- , Helen Cox
- , Christopher Davis
- , Oslem Dincarslan
- , Chris Dunn
- , Philip Dyer
- , Angela Elliott
- , Anthony Evans
- , Lorna Finch
- , Lewis W. S. Fisher
- , Terry Foster
- , Isabel Garcia-Dorival
- , William Greenhalf
- , Philip Gunning
- , Catherine Hartley
- , Rebecca L. Jensen

- , Christopher B. Jones
- , Trevor R. Jones
- , Shadia Khandaker
- , Katharine King
- , Robyn T. Kiy
- , Chrysa Koukorava
- , Annette Lake
- , Suzannah Lant
- , Diane Latawiec
- , Lara Lavelle-Langham
- , Daniella Lefteri
- , Lauren Lett
- , Lucia A. Livoti
- , Maria Mancini
- , Sarah McDonald
- , Laurence McEvoy
- , John McLauchlan
- , Soeren Metelmann
- , Nahida S. Miah
- , Joanna Middleton
- , Joyce Mitchell
- , Shona C. Moore
- , Ellen G. Murphy
- , Rebekah Penrice-Randal
- , Jack Pilgrim
- , Tessa Prince
- , Will Reynolds
- , P. Matthew Ridley
- , Debby Sales
- , Victoria E. Shaw
- , Rebecca K. Shears
- , Benjamin Small
- , Krishanthi S. Subramaniam
- , Agnieszka Szemiel
- , Aislynn Taggart
- , Jolanta Tanianis-Hughes
- , Jordan Thomas

- , Erwan Trochu
- , Libby van Tonder
- , Eve Wilcock
- , J. Eunice Zhang
- , Lisa Flaherty
- , Nicole Maziere
- , Emily Cass
- , Alejandra Doce Carracedo
- , Nicola Carlucci
- , Anthony Holmes
- , Hannah Massey
- , Lee Murphy
- , Nicola Wrobel
- , Sarah McCafferty
- , Kirstie Morrice
- & Alan MacLean

- **Admin team members**

- Ruth Armstrong
- , J. Kenneth Baillie
- , Ceilia Boz
- , Adam Brown
- , Richard Clark
- , Sara Clohisey
- , Audrey Coutts
- , Louise Cullum
- , Nicky Day
- , Lorna Donnelly
- , Esther Duncan
- , Angie Fawkes
- , Paul Finernan
- , Max Head Fourman
- , James Furniss
- , Tammy Gilchrist
- , Ailsa Golightly
- , Katarzyna Hafezi

- , Ross Hendry
- , Andy Law
- , Dawn Law
- , Rachel Law
- , Sarah Law
- , Louise Macgillivray
- , Alan Maclean
- , Hanning Mal
- , Sarah McCafferty
- , Ellie Mcmaster
- , Jen Meikle
- , Shona C. Moore
- , Kirstie Morrice
- , Lee Murphy
- , Wilna Oosthuyzen
- , Nicholas Parkinson
- , Trevor Paterson
- , Andrew Stenhouse
- , Maaike Swets
- , Helen Szoor-McElhinney
- , Filip Taneski
- , Lance C. W. Turtle
- , Tony Wackett
- , Mairi Ward
- , Jane Weaver
- , Nicola Wrobel
- , Marie Zechner
- , Judy Coyle
- , Bernadette Gallagher
- , Rebecca Lidstone-Scott
- , Debbie Hamilton
- , Katherine Schon
- , Anita Furlong
- , Heather Biggs
- , Fiona Griffiths
- , Eleanor Andrews
- , Kathy Brickell

- , Michelle Smyth
  - , Lorna Murphy
  - , Gail Carson
  - , Hayley Hardwick
  - & Chloe Donohue
- **COVID-19 HGI corresponding authors**
    - Benjamin M. Neale
    - , Mark Daly
    - & Andrea Ganna

## Contributions

Author contributions are provided within the author list.

## Corresponding authors

Correspondence to [Benjamin M. Neale](#), [Mark Daly](#) or [Andrea Ganna](#).

## Ethics declarations

## Competing interests

A full list of competing interests is supplied as Supplementary Table 13.

## Additional information

**Peer review information** *Nature* thanks Samira Asgari, Paul McLaren and Neneh Sallah for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Analytical summary of the COVID-19 HGI meta-analysis.

Using the analytical plan set by the COVID-19 HGI, each individual study runs their analyses and uploads the results to the Initiative, who then runs the meta-analysis. There are three main analyses that each study can contribute summary statistics to: critically ill COVID-19, hospitalized COVID-19 and reported SARS-CoV-2 infection. The phenotypic criteria used to define cases are listed in the dark grey boxes, along with the numbers of cases ( $N$ ) included in the final all-ancestries meta-analysis. Controls were defined in the same way across all three analyses as everybody that is not a case—for example, population controls (light grey box). Sensitivity analyses—not reported in this extended data figure—also included mild and/or asymptomatic cases of COVID-19 as control individuals. Sample number ( $N$ ) of control individuals differed between the analyses due to the difference in the number of studies contributing data to these.

### Extended Data Fig. 2 Projection of contributing studies samples into the same PC space.

We asked participating studies to perform a PC projection using the 1000 Genomes Project and Human Genome Diversity Project as a reference, with a common set of variants. For each panel (except for the reference), coloured points correspond to contributed samples from each cohort, whereas grey points correspond to the reference samples from the 1000 Genomes Project. Colour represents a genetic population that each cohort specified. As 23andMe, Genomics England 100,000 Genomes Project (GenomicsEngland100kgp), and Million Veterans Program (MVP) only submitted PCA images, we overlaid their submitted transparent images using the same coordinates, instead of directly plotting them. Populations are defined as African (AFR), admixed American (AMR), East Asian (EAS), European (EUR), Middle Eastern (MID) and South Asian (SAS), Oceanian (OCE).

## Extended Data Fig. 3 Locus-zoom plots of the 3p21.31 region for reported SARS-CoV-2 infection.

**a**, A standard plot without exclusion. Here, the severity lead variant rs10490770 (chr. 3: 45823240T:C) is shown as a lead variant. **b**, Additional independent susceptibility signal(s) after excluding variants with  $r^2 > 0.05$  with rs10490770. The susceptibility lead variant rs2271616 (chr. 3: 45796521G:T) is highlighted.

## Extended Data Fig. 4 Genome-wide meta-analysis association results for critical illness due to COVID-19.

The locus on chromosome 6 is the HLA locus, which was removed from the list of reported loci in Supplementary Table 2 due to the high heterogeneity in effect size estimated between studies included in the analysis. The locus on chromosome 7 was also not reported in Supplementary Table 2 due to missingness across studies—that is, the high number of studies in the meta-analysis that did not report summary statistics for this region. There are two association peaks on chromosome 19.

## Extended Data Fig. 5 Sensitivity analyses for overlapping controls in genomiCC and UK Biobank.

Comparison of the beta effect sizes (top) and unadjusted  $P$  values (bottom) of the 13 lead variants, using data from the COVID-19 critical illness meta-analysis in all the cohorts to leaving out genomiCC (cases,  $n = 4,354$ ; controls,  $n = 1,474,655$ ; total,  $n = 1,479,009$ ), leaving out the UK Biobank (UKBB; cases,  $n = 5,870$ ; controls,  $n = 1,155,203$ ; total,  $n = 1,161,073$ ) and leaving out both genomiCC and UK Biobank (cases,  $n = 4,045$ ; controls,  $n = 1,146,078$ ; total,  $n = 1,150,123$ ) (from left to right, respectively). Top, dots and grey bars represent the beta effect size estimates  $\pm$  standard error from the corresponding GWAS meta-analysis. Bottom, dots represent two-sided  $P$  values from the corresponding GWAS meta-analysis. Filled dots indicate variants that showed genome-wide significance in the full meta-analysis of critical illness due to COVID-19, and empty dots represent variants that were not significant for critical illness but were significant for

either hospitalization due to COVID-19 or reported SARS-CoV-2 infection. Red dots represent variants that showed genome-wide significance in the leave-one-out analysis for genomiCC, UK Biobank or genomiCC and UK Biobank.

**Extended Data Fig. 6 Comparison of  $\chi^2$  statistics and  $r^2$  values to the lead variant in the 3p21.31 region.**

**a–c**, Data are shown for critical illness (**a**), hospitalization (**b**) and reported SARS-CoV-2 infection (**c**). The left blue peak in **c**, which is uncorrelated with the lead variants in the region, indicates that there are independent signals.

**Extended Data Fig. 7 Comparison of the effect sizes of lead variants between pairs of COVID-19 meta-analyses.**

Comparison of effect sizes for the nine variants associated with severity of COVID-19 disease. **a**, Comparing hospitalized cases of COVID-19 versus population controls ( $n = 10,428$  cases and  $n = 1,483,270$  controls) and critically ill cases of COVID-19 versus population controls ( $n = 6,179$  cases and  $n = 1,483,780$  controls). **b**, Hospitalized cases of COVID-19 versus population controls ( $n = 5,806$  cases and  $n = 1,144,263$  controls) and hospitalized cases of COVID-19 versus non-hospitalized cases of COVID-19 ( $n = 5,773$  cases and  $n = 15,497$  controls). Sample sizes for hospitalized cases of COVID-19 versus population controls differ between **a** and **b** due to differences in the sampling of studies selected for the analysis. This selection included all studies that were able to contribute data to the respective analyses that the data were compared to (shown on the  $y$  axis) in each panel. Dots represent the effect size beta estimates, bars represent the 95% confidence interval of the estimates. Effect size estimates and  $P$  values for heterogeneity tests (Cochran's  $Q$ , two-tailed test) are reported in Supplementary Table 3.

**Extended Data Fig. 8 PheWAS for genome-wide significant lead variants.**

Selected phenotypes associated with genome-wide significant COVID-19 variants (see Supplementary Table [6](#) for a complete list). We report those associations for which a lead variant from a previous GWAS result was in high LD ( $r^2 > 0.8$ ) with the index COVID-19 variants. The colour represents the z-scores of correlated risk increasing alleles for the trait. The total number of associations for each COVID-19 variant is highlighted in the grey box.

### **Extended Data Fig. 9 Genetic correlation with COVID-19 phenotypes.**

Each column shows the genetic correlation results for the three COVID-19 phenotypes (European-ancestry analyses only): critical illness, hospitalization and reported SARS-CoV-2 infection. The traits that the genetic correlation is run against are listed on the left. Significant correlations (FDR  $< 0.05$ ) are shown with their 95% confidence intervals in red, nominally significant correlations ( $P < 0.05$ ) are in black and non-significant correlations are in grey. Two-sided  $P$  values were calculated using LDSC for genetic correlations and exact estimates, unadjusted standard errors and two-sided  $P$  values are available in Supplementary Table [11](#).

### **Extended Data Fig. 10 Mendelian randomization sensitivity analyses.**

Genetic correlations and Forest plots displaying the causal estimates for each of the sensitivity analyses used in the Mendelian randomization analysis for trait pairs that were significant at an FDR of 5%. Two-sided  $P$  values were estimated using IVW, WME, WMBE and MR-PRESSO analyses. RBC, red blood cell count.

## **Supplementary information**

### **Supplementary Information**

This Supplementary Information file contains the following sections: New and replicated loci from COVID-19 HGI meta-analyses; Additional independent susceptibility signals at the 3p21.31 locus ; Sensitivity analysis for use of population controls; Sensitivity analysis for overlapping samples between cohorts in Mendelian randomization analyses; Supplementary discussion on study limitations; Supplementary References; and titles and summaries for Supplementary Tables 1-13 (see Excel file for Supplementary Tables).

## Reporting Summary

### Supplementary Figure 1

Quantile-quantile plots for GWAS from all individual studies that contributed data. QQ-plots showing the expected  $-\log_{10}(P\text{-values})$  on the x-axis and the observed unadjusted  $P$ -values values from two-tailed inverse variance weighted meta-analysis on the y-axis (red line showing no deviation from the expected) for each study contributing data to the analyses. Sample size of cases and controls is listed for each study in the plot title, as well as the median lambda value.

### Supplementary Figure 2

LozusZoom plots to visualise the meta-analysis results at the loci passing genome-wide significance. For each genome-wide significant locus in three meta-analyses: meta-analysis of critical illness, hospitalization, and reported infection, we showed 1) a manhattan plot of each locus where a color represents a weighted-average  $r^2$  value (see Methods) to a lead variant (unadjusted  $P$ -values from the two-tailed inverse variance weighted meta-analysis); 2)  $r^2$  values to a lead variant across gnomAD v2 populations, i.e., African/African-American (AFR), Latino/Admixed American (AMR), Ashkenazi Jewish (ASJ), East Asian (EAS), Estonian (EST), Finnish (FIN), Non-Finish Europeans (NFE), North-Western Europeans (NWE), and Southern Europeans (SEU); 3) genes at a locus; and 4) genes prioritized by each gene prioritization metric where a size of circles represents a rank in each metric. Note that the COVID-19 lead variants were chosen across all

the meta-analyses (Supplementary Table 2; see Methods) and were not necessarily a variant with the most significant *P*-value from each inverse variance weighted meta-analysis.

### **Supplementary Figure 3**

Scatter and funnel plots for each exposure - COVID-19 outcome pair. Scatter plots show the exposure variant effect size against the COVID-19 outcome variant effect size and corresponding standard errors. Funnel plots show the Mendelian randomization (MR) causal estimates for each variant against their precision, with asymmetry in the plot indicating potential violations of the assumptions of MR. Regression lines show the corresponding causal estimates fixed effect inverse-weighted (IVW, red-solid line); MR-Egger regression (blue-dashed); Weighted median estimator (WME, green-dashed); weighted mode based estimator (WMBE, purple-dashed); and Mendelian Randomization Pleiotropy RESidual Sum and Outlier corrected (MR-PRESSO, orange-dashed). Variants highlighted in red were flagged as outliers by MR-PRESSO.

### **Supplementary Tables**

This file contains Supplementary Tables 1-13; see main Supplementary Information PDF for table titles and summaries.

### **Supplementary Information**

This file contains the full authorship for the Covid-19 Host Genetics Initiative.

### **Peer Review File**

### **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## [Reprints and Permissions](#)

## About this article

### Cite this article

COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).  
<https://doi.org/10.1038/s41586-021-03767-x>

- Received: 02 March 2021
- Accepted: 23 June 2021
- Published: 08 July 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-03767-x>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [Update on human genetic susceptibility to COVID-19: susceptibility to virus and response](#)
  - Vito Luigi Colona
  - Vasilis Vasiliou
  - Juergen K. V. Reichardt

*Human Genomics* (2021)
- [Estimating causal effects of atherogenic lipid-related traits on COVID-19 susceptibility and severity using a two-sample Mendelian randomization approach](#)
  - Masahiro Yoshikawa
  - Kensuke Asaba
  - Tomohiro Nakayama

*BMC Medical Genomics* (2021)
- [Lung microbiota and COVID-19 severity](#)
  - Robert P. Dickson

*Nature Microbiology* (2021)
- [Metabolic disorders, COVID-19 and vaccine-breakthrough infections](#)
  - Norbert Stefan

*Nature Reviews Endocrinology* (2021)

- **Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus**

- Damien J. Downes
- Amy R. Cross
- Jim R. Hughes

*Nature Genetics* (2021)

### **Human genetic variants identified that affect COVID susceptibility and severity**

- Samira Asgari
- Lionel A. Pousaz

News & Views 08 Jul 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03767-x>

- Article
- [Published: 08 December 2021](#)

# Megastudies improve the impact of applied behavioural science

- [Katherine L. Milkman](#) ORCID: [orcid.org/0000-0002-9706-4830<sup>1</sup>](https://orcid.org/0000-0002-9706-4830),
- [Dena Gromet<sup>2</sup>](#),
- [Hung Ho<sup>1</sup>](#) nAff26,
- [Joseph S. Kay<sup>2</sup>](#),
- [Timothy W. Lee<sup>2</sup>](#) nAff27,
- [Pepi Pandiloski<sup>3</sup>](#),
- [Yeji Park<sup>4</sup>](#),
- [Aneesh Rai](#) ORCID: [orcid.org/0000-0001-9633-2568<sup>1</sup>](https://orcid.org/0000-0001-9633-2568),
- [Max Bazerman<sup>5</sup>](#),
- [John Beshears<sup>5</sup>](#),
- [Lauri Bonacorsi<sup>6</sup>](#),
- [Colin Camerer<sup>7</sup>](#),
- [Edward Chang](#) ORCID: [orcid.org/0000-0003-0502-2957<sup>5</sup>](https://orcid.org/0000-0003-0502-2957),
- [Gretchen Chapman<sup>8</sup>](#),
- [Robert Cialdini<sup>9</sup>](#),
- [Hengchen Dai](#) ORCID: [orcid.org/0000-0001-7640-6558<sup>10</sup>](https://orcid.org/0000-0001-7640-6558),
- [Lauren Eskreis-Winkler<sup>11</sup>](#),
- [Ayelet Fishbach<sup>11</sup>](#),
- [James J. Gross](#) ORCID: [orcid.org/0000-0003-3624-3090<sup>12</sup>](https://orcid.org/0000-0003-3624-3090),
- [Samantha Horn](#) ORCID: [orcid.org/0000-0003-2716-5860<sup>8</sup>](https://orcid.org/0000-0003-2716-5860),
- [Alexa Hubbard<sup>13</sup>](#),
- [Steven J. Jones<sup>14</sup>](#),
- [Dean Karlan](#) ORCID: [orcid.org/0000-0003-0164-1398<sup>15</sup>](https://orcid.org/0000-0003-0164-1398),
- [Tim Kautz<sup>16</sup>](#),

- [Erika Kirgios](#) ORCID: [orcid.org/0000-0002-7891-0499<sup>1</sup>](https://orcid.org/0000-0002-7891-0499),
- [Joowon Klusowski<sup>17</sup>](#),
- [Ariella Kristal<sup>18</sup>](#),
- [Rahul Ladhania](#) ORCID: [orcid.org/0000-0002-7902-7681<sup>19</sup>](https://orcid.org/0000-0002-7902-7681),
- [George Loewenstein](#) ORCID: [orcid.org/0000-0003-2790-0474<sup>8</sup>](https://orcid.org/0000-0003-2790-0474),
- [Jens Ludwig<sup>3</sup>](#),
- [Barbara Mellers<sup>17</sup>](#),
- [Sendhil Mullainathan<sup>11</sup>](#),
- [Silvia Saccardo<sup>8</sup>](#),
- [Jann Spiess<sup>20</sup>](#),
- [Gaurav Suri<sup>21</sup>](#),
- [Joachim H. Talloen<sup>8</sup>](#),
- [Jamie Taxer<sup>12</sup>](#),
- [Yaakov Trope<sup>13</sup>](#),
- [Lyle Ungar<sup>22</sup>](#),
- [Kevin G. Volpp<sup>23</sup>](#),
- [Ashley Whillans<sup>5</sup>](#),
- [Jonathan Zinman<sup>24</sup>](#) &
- [Angela L. Duckworth](#) ORCID: [orcid.org/0000-0002-2644-3729<sup>1,25</sup>](https://orcid.org/0000-0002-2644-3729)

[Nature](#) volume 600, pages 478–483 (2021)

- 9966 Accesses
- 1 Citations
- 692 Altmetric
- [Metrics details](#)

## Subjects

- [Decision making](#)
- [Human behaviour](#)

# Abstract

Policy-makers are increasingly turning to behavioural science for insights about how to improve citizens' decisions and outcomes<sup>1</sup>. Typically, different scientists test different intervention ideas in different samples using different outcomes over different time intervals<sup>2</sup>. The lack of comparability of such individual investigations limits their potential to inform policy. Here, to address this limitation and accelerate the pace of discovery, we introduce the megastudy—a massive field experiment in which the effects of many different interventions are compared in the same population on the same objectively measured outcome for the same duration. In a megastudy targeting physical exercise among 61,293 members of an American fitness chain, 30 scientists from 15 different US universities worked in small independent teams to design a total of 54 different four-week digital programmes (or interventions) encouraging exercise. We show that 45% of these interventions significantly increased weekly gym visits by 9% to 27%; the top-performing intervention offered microrewards for returning to the gym after a missed workout. Only 8% of interventions induced behaviour change that was significant and measurable after the four-week intervention. Conditioning on the 45% of interventions that increased exercise during the intervention, we detected carry-over effects that were proportionally similar to those measured in previous research<sup>3,4,5,6</sup>. Forecasts by impartial judges failed to predict which interventions would be most effective, underscoring the value of testing many ideas at once and, therefore, the potential for megastudies to improve the evidentiary value of behavioural science.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

### Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

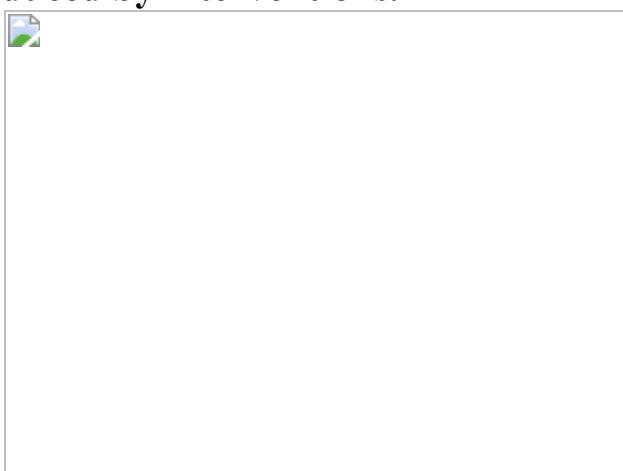
### Rent or Buy

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Measured versus predicted changes in weekly gym visits induced by interventions.**



### **Data availability**

The data analysed in this paper were provided by 24 Hour Fitness and we have their legal permission to share the deidentified data. We have therefore made deidentified data available at [https://osf.io/9av87/?view\\_only=8bb9282111c24f81a19c2237e7d7eba3](https://osf.io/9av87/?view_only=8bb9282111c24f81a19c2237e7d7eba3). Furthermore, tables of all of the preregistration links for each of the substudies with the interventions and the prediction studies are available in Supplementary Tables 2 and 30.

## Code availability

The code to replicate the analyses and figures in the paper and **Supplementary Information** is available online ([https://osf.io/9av87/?view\\_only=8bb9282111c24f81a19c2237e7d7eba3](https://osf.io/9av87/?view_only=8bb9282111c24f81a19c2237e7d7eba3)).

## References

1. 1.  
*Behavioural Insights and Public Policy: Lessons from Around the World* (OECD Publishing, 2017).
2. 2.  
Benartzi, S. et al. Should governments invest more in nudging? *Psychol. Sci.* **28**, 1041–1055 (2017).
3. 3.  
Charness, G. & Gneezy, U. Incentives to EXercise. *Econometrica* **77**, 909–931 (2009).
4. 4.  
Acland, D. & Levy, M. R. Naiveté, projection bias, and habit formation in gym attendance. *Manage. Sci.* **61**, 146–160 (2015).
5. 5.

Royer, H., Stehr, M. & Sydnor, J. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a Fortune-500 company. *Am. Econ. J. Appl. Econ.* **7**, 51–84 (2015).

6. 6.

Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R. & Wisdom, J. Creating exercise habits using incentives: the tradeoff between flexibility and routinization. *Manage. Sci.* **67**, 4139–4171 (2020).

7. 7.

DellaVigna, S. & Linos, E. *RCTs to Scale: Comprehensive Evidence from Two Nudge Units* 65 (National Bureau of Economic Research, 2020).

8. 8.

DellaVigna, S. & Pope, D. What motivates effort? Evidence and expert forecasts. *Rev. Econ. Stud.* **85**, 1029–1069 (2018).

9. 9.

DellaVigna, S. & Pope, D. Predicting experimental results: who knows what? *J. Polit. Econ.* **126**, 2410–2456 (2018).

10. 10.

DellaVigna, S., Pope, D. & Vivaldi, E. Predict science to improve science. *Science* **366**, 428–429 (2019).

11. 11.

Kristal, A. S. & Whillans, A. V. What we can learn from five naturalistic field experiments that failed to shift commuter behaviour. *Nat. Hum. Behav.* **4**, 169–176 (2020).

12. 12.

Donoho, D. 50 years of data science. *J. Comput. Graph. Stat.* **26**, 745–766 (2017).

13. 13.

Liberman, M. Fred Jelinek. *Comput. Linguist.* **36**, 595–599 (2010).

14. 14.

Lai, C. K. et al. Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* **143**, 1765–1785 (2014).

15. 15.

Lai, C. K. et al. Reducing implicit racial preferences: II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* **145**, 1001–1016 (2016).

16. 16.

Mellers, B. et al. Psychological strategies for winning a geopolitical forecasting tournament. *Psychol. Sci.* **25**, 1106–1115 (2014).

17. 17.

Open Science Collaboration Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).

18. 18.

Milkman, K. L., Minson, J. A. & Volpp, K. G. M. Holding the hunger games hostage at the gym: an evaluation of temptation bundling. *Manage. Sci.* **60**, 283–299 (2014).

19. 19.

Ward, B. W., Clarke, T. C., Nugent, C. N. & Schiller, J. S. *Early Release of Selected Estimates Based on Data From the 2015 National*

*Health Interview Survey* 120 (National Center for Health Statistics, 2015).

20. 20.

Lee, I.-M. et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* **380**, 219–229 (2012).

21. 21.

Gollwitzer, P. M. Implementation intentions: strong effects of simple plans. *Am. Psychol.* **54**, 493–503 (1999).

22. 22.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. Using implementation intentions prompts to enhance influenza vaccination rates. *Proc. Natl Acad. Sci. USA* **108**, 10415–10420 (2011).

23. 23.

Rogers, T., Milkman, K. L., John, L. K. & Norton, M. I. Beyond good intentions: prompting people to make plans improves follow-through on important tasks. *Behav. Sci. Pol.* **1**, 33–41 (2015).

24. 24.

Karlan, D., McConnell, M., Mullainathan, S. & Zinman, J. Getting to the top of mind: how reminders increase saving. *Manage. Sci.* **62**, 3393–3411 (2016).

25. 25.

Homonoff, T. A. Can small incentives have large effects? The impact of taxes versus bonuses on disposable bag use. *Am. Econ. J. Econ. Pol.* **10**, 177–210 (2018).

26. 26.

Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).

27. 27.

Allcott, H. Social norms and energy conservation. *J. Publ. Econ.* **95**, 1082–1095 (2011).

28. 28.

Chapman, G. B., Li, M., Colby, H. & Yoon, H. Opting in vs opting out of influenza vaccination. *JAMA* **304**, 43–44 (2010).

29. 29.

Milkman, K. L., et al. A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor’s appointment. *Proc. Natl Acad. Sci. USA* **118**, e2101165118 (2021).

30. 30.

Lee, M. R. & Shen, M. Winner’s curse: bias estimation for total effects of features in online controlled experiments. In *Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 491–499 (ACM, 2018).

31. 31.

White, H. *Asymptotic Theory for Econometricians* (Elsevier, 1984).

32. 32.

Dubner, S. J. How goes the behavior-change revolution? (Ep. 382). *Freakonomics* <https://freakonomics.com/podcast/live-philadelphia/> (2019).

# Acknowledgements

Support for this research was provided in part by the Robert Wood Johnson Foundation, the AKO Foundation, J. Alexander, M. J. Leder, W. G. Lichtenstein, the Pershing Square Fund for Research on the Foundations of Human Behavior from Harvard University and by Roybal Center grants (P30AG034546 and 5P30AG034532) from the National Institute on Aging. The views expressed here do not necessarily reflect the views of any of these individuals or entities. We thank 24 Hour Fitness for partnering with the Behavior Change for Good Initiative at the University of Pennsylvania to make this research possible.

# Author information

## Author notes

### 1. Hung Ho

Present address: Department of Marketing, Booth School of Business, University of Chicago, Chicago, IL, USA

### 2. Timothy W. Lee

Present address: McCormick School of Engineering, Northwestern University, Evanston, IL, USA

# Affiliations

### 1. Department of Operations, Information and Decisions, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Katherine L. Milkman, Hung Ho, Aneesh Rai, Erika Kirgios & Angela L. Duckworth

### 2. Behavior Change for Good Initiative, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Dena Gromet, Joseph S. Kay & Timothy W. Lee

3. Harris School of Public Policy, University of Chicago, Chicago, IL, USA

Pepi Pandiloski & Jens Ludwig

4. Department of Psychology, Princeton University, Princeton, NJ, USA

Yeqi Park

5. Department of Negotiation, Organizations & Markets, Harvard Business School, Harvard University, Boston, MA, USA

Max Bazerman, John Beshears, Edward Chang & Ashley Whillans

6. Pritzker School of Law, Northwestern University, Chicago, IL, USA

Lauri Bonacorsi

7. Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

Colin Camerer

8. Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA

Gretchen Chapman, Samantha Horn, George Loewenstein, Silvia Saccardo & Joachim H. Talloen

9. Department of Psychology, Arizona State University, Tempe, AZ, USA

Robert Cialdini

10. Department of Management and Organizations, Anderson School of Management, University of California Los Angeles, Los Angeles, CA, USA

Hengchen Dai

11. Department of Behavioral Science, Booth School of Business,  
University of Chicago, Chicago, IL, USA

Lauren Eskreis-Winkler, Ayelet Fishbach & Sendhil Mullainathan

12. Department of Psychology, Stanford University, Stanford, CA, USA

James J. Gross & Jamie Taxer

13. Department of Psychology, New York University, New York, NY,  
USA

Alexa Hubbard & Yaacov Trope

14. Department of Psychology, Rutgers University, New Brunswick, NJ,  
USA

Steven J. Jones

15. Department of Finance, Kellogg School of Management, Northwestern  
University, Evanston, IL, USA

Dean Karlan

16. Mathematica, Princeton, NJ, USA

Tim Kautz

17. Department of Marketing, The Wharton School, University of  
Pennsylvania, Philadelphia, PA, USA

Joowon Klusowski & Barbara Mellers

18. Department of Organizational Behavior, Harvard Business School,  
Harvard University, Boston, MA, USA

Ariella Kristal

19. Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI, USA

Rahul Ladhania

20. Department of Operations, Information & Technology, Stanford Graduate School of Business, Stanford, CA, USA

Jann Spiess

21. Department of Psychology, San Francisco State University, San Francisco, CA, USA

Gaurav Suri

22. Department of Computer and Information Sciences, University of Pennsylvania, Philadelphia, PA, USA

Lyle Ungar

23. Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Kevin G. Volpp

24. Department of Economics, Dartmouth College, Hanover, NH, USA

Jonathan Zinman

25. Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Angela L. Duckworth

## **Contributions**

K.L.M., D.G., A.R., M.B., J.B., L.B., E.C., G.C., R.C., H.D., L.E.-W., A.F., J.J.G., S.H., A.H., S.J.J., D.K., E.K., J.K., A.K., G.L., B.M., S.M., S.S., G.S., J.H.T., J.T., Y.T., L.U., K.G.V., A.W., J.Z. and A.L.D. designed the

research. K.L.M., D.G., J.S.K., P.P., Y.P., A.L.D. and A.R. performed the research. H.H., T.W.L., P.P. and Y.P. analysed the data. K.L.M. and A.L.D wrote the paper. D.G., H.H., J.S.K., T.W.L., P.P., Y.P., A.R., M.B., J.B., C.C., G.C., H.D., A.F., J.J.G., D.K., T.K., E.K., J.K., R.L., J.L., B.M., S.M., S.S., J.S., A.W. and J.Z. provided feedback on the paper. K.L.M., D.G., J.S.K., T.K., R.L. and S.M. supervised data analysis. K.L.M., D.G., H.H., J.S.K. and T.W.L. prepared the Supplementary Information.

## Corresponding authors

Correspondence to [Katherine L. Milkman](#) or [Angela L. Duckworth](#).

## Ethics declarations

### Competing interests

The authors declare no competing interests. The authors did not receive commercial benefits from the fitness chain or speaking/consulting fees related to any of the interventions presented here.

## Additional information

**Peer review information** *Nature* thanks Charles Shearer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

[Extended Data Fig. 1 Measured vs. predicted change in likelihood of gym visit in a given week.](#)

The measured change (blue) vs. change predicted by third-party observers (gold) in whether participants visited the gym that was induced by each of our megastudy's 53 experimental conditions compared to a Placebo Control condition during a four-week intervention period is depicted here. Error bars represent 95% confidence intervals. See Extended Data Table [7](#) for complete OLS regression results graphed here in blue, Supplementary Information [11](#) for more details about the prediction data graphed here in gold, and Supplementary Table [1](#) for full descriptions of each treatment condition in our megastudy. Sample weights were included in the pooled third-party prediction data to ensure equal weighting of each of our three participant samples (professors, practitioners and prolific respondents). The superscripts a–e denote the different incentive amounts offered in different versions of the bonus for returning after missed workouts, higher incentives and rigidity rewarded conditions, which are described in Supplementary Table 1. In conditions with the same name, superscripts that come earlier in the alphabet indicate larger incentives.

**Extended Data Table 1 Regression-estimated effects of each experimental condition on whether participants visited the gym in a given week during the four-week intervention period relative to the Planning, Reminders and Micro-Incentives to Exercise condition**

**Extended Data Table 2 Regression-estimated effects of each experimental condition on whether participants visited the gym in a given week during the four-week post-intervention period relative to the Placebo Control condition**

**Extended Data Table 3 The percentage of other conditions that each experimental condition outperformed for our dependent variable measuring whether participants visited the gym in a given week at  $p < .05$  during the four-week intervention period**

**Extended Data Table 4 Participants' mean age (in years), gender, length of gym membership (in weeks), and mean weekly gym visits in the four-week pre-intervention period across the 54 study conditions**

**Extended Data Table 5 Percentage of significant p-values and absolute difference in coefficients from pairwise comparisons of the 54 study conditions in our megastudy on each variable listed (alpha = .05)**

**Extended Data Table 6 Regression-estimated effects of each experimental condition on total weekly gym visits during the four-week**

**intervention period relative to the Placebo Control condition**

**Extended Data Table 7 Regression-estimated effects of each experimental condition on whether participants visited the gym in a given week during the four-week intervention period relative to the Placebo Control condition**

**Extended Data Table 8 Regression-estimated effects of each experimental condition on total weekly gym visits during the four-week intervention period relative to the Planning, Reminders, and Micro-Incentives to Exercise condition**

**Extended Data Table 9 Regression-estimated effects of each experimental condition on total weekly gym visits during the four-week post-intervention period relative to the Placebo Control condition**

## **Supplementary information**

### **Supplementary Information**

Supplementary Figures, Tables and Notes.

### **Reporting Summary**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Milkman, K.L., Gromet, D., Ho, H. *et al.* Megastudies improve the impact of applied behavioural science. *Nature* **600**, 478–483 (2021).

<https://doi.org/10.1038/s41586-021-04128-4>

- Received: 22 October 2020

- Accepted: 13 October 2021
- Published: 08 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04128-4>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*Benefits of megastudies for testing behavioural interventions\*\*](#)
  - Heather Royer

*Nature* (2021)

## [\*\*How 'megastudies' are changing behavioural science\*\*](#)

- Benjamin Thompson
- Shamini Bundell

Nature Podcast 08 Dec 2021

## [\*\*Benefits of megastudies for testing behavioural interventions\*\*](#)

- Heather Royer

## News & Views 08 Dec 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04128-4>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 10 November 2021](#)

# Cognitive control persistently enhances hippocampal information processing

- [Ain Chung<sup>1</sup>](#),
- [Claudia Jou<sup>2</sup>](#),
- [Alejandro Grau-Perales<sup>1</sup>](#),
- [Elliott R. J. Levy<sup>1</sup>](#),
- [Dino Dvorak<sup>1</sup>](#),
- [Nida Hussain<sup>1</sup>](#) &
- [André A. Fenton](#)    [ORCID: orcid.org/0000-0002-5063-1156<sup>1,3</sup>](#)

[Nature](#) volume 600, pages 484–488 (2021)

- 7012 Accesses
- 380 Altmetric
- [Metrics details](#)

## Subjects

- [Hippocampus](#)
- [Inhibition–excitation balance](#)
- [Long-term memory](#)
- [Spatial memory](#)

## Abstract

Could learning that uses cognitive control to judiciously use relevant information while ignoring distractions generally improve brain function, beyond forming explicit memories? According to a neuroplasticity hypothesis for how some cognitive behavioural therapies are effective, cognitive control training (CCT) changes neural circuit information processing<sup>1,2,3</sup>. Here we investigated whether CCT persistently alters hippocampal neural circuit function. We show that mice learned and remembered a conditioned place avoidance during CCT that required ignoring irrelevant locations of shock. CCT facilitated learning new tasks in novel environments for several weeks, relative to unconditioned controls and control mice that avoided the same place during reduced distraction. CCT rapidly changes entorhinal cortex-to-dentate gyrus synaptic circuit function, resulting in an excitatory–inhibitory subcircuit change that persists for months. CCT increases inhibition that attenuates the dentate response to medial entorhinal cortical input, and through disinhibition, potentiates the response to strong inputs, pointing to overall signal-to-noise enhancement. These neurobiological findings support the neuroplasticity hypothesis that, as well as storing item–event associations, CCT persistently optimizes neural circuit information processing.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

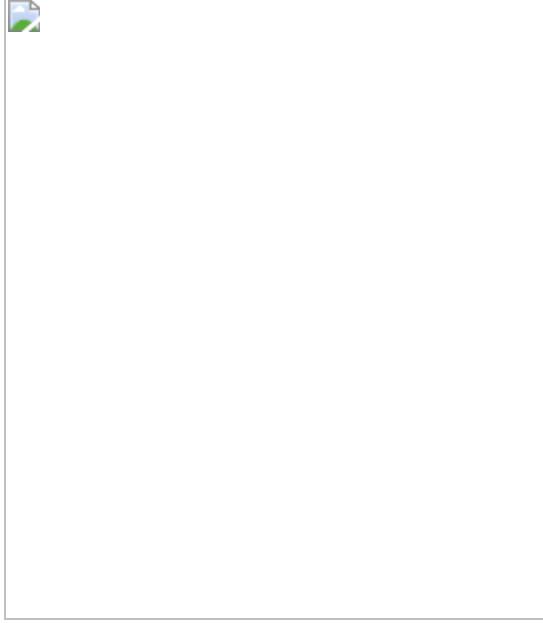
[Rent or Buy](#)

All prices are NET prices.

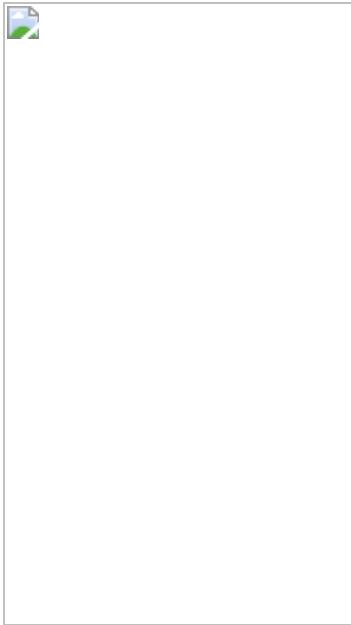
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

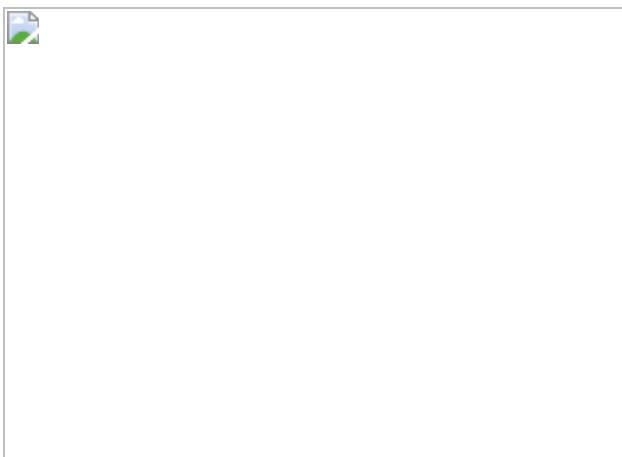
**Fig. 1: Cognitive control of spatial information processing in CA1.**



**Fig. 2: CCT facilitates subsequent learning.**



**Fig. 3: Cognitive control training persistently changes entorhinal-dentate circuit function.**



**Fig. 4: Cognitive control training persistently changes inhibitory entorhinal-dentate circuit function.**



## Data availability

The datasets are available from the corresponding author upon reasonable request. Source data are available at  
[https://github.com/aafenton/Chung\\_etal\\_LearningToLearnData.git](https://github.com/aafenton/Chung_etal_LearningToLearnData.git). Source data are provided with this paper.

## Code availability

Code used for data processing and analysis is available from the corresponding author upon reasonable request at  
[https://github.com/aafenton/Chung\\_etal\\_LearningToLearnCode.git](https://github.com/aafenton/Chung_etal_LearningToLearnCode.git).

## References

1. 1.

Subramaniam, K. et al. Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. *Neuron* **73**, 842–853 (2012).

2. 2.

Anguera, J. A. et al. Video game training enhances cognitive control in older adults. *Nature* **501**, 97–101 (2013).

3. 3.

Mishra, J., de Villers-Sidani, E., Merzenich, M. & Gazzaley, A. Adaptive training diminishes distractibility in aging across species. *Neuron* **84**, 1091–1103 (2014).

4. 4.

Thomas, C. & Baker, C. I. Teaching an adult brain new tricks: a critical review of evidence for training-dependent structural plasticity in humans. *Neuroimage* **73**, 225–236 (2013).

5. 5.

Alonso, A., van der Meij, J., Tse, D. & Genzel, L. Naïve to expert: Considering the role of previous knowledge in memory. *Brain. Neurosci. Adv.* **4**, 2398212820948686 (2020).

6. 6.

Yeager, D. S. et al. A national experiment reveals where a growth mindset improves achievement. *Nature* **573**, 364–369 (2019).

7. 7.

Bavelier, D., Green, C. S., Pouget, A. & Schrater, P. Brain plasticity through the life span: learning to learn and action video games. *Annu. Rev. Neurosci.* **35**, 391–416 (2012).

8. 8.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.* **56**, 51–65 (1949).

9. 9.

Brewin, C. R. Theoretical foundations of cognitive-behavior therapy for anxiety and depression. *Annu. Rev. Psychol.* **47**, 33–57 (1996).

10. 10.

de Villers-Sidani, E. et al. Recovery of functional and structural age-related changes in the rat primary auditory cortex with operant training. *Proc. Natl Acad. Sci. USA* **107**, 13900–13905 (2010).

11. 11.

Pavlovsky, A., Wallace, E., Fenton, A. A. & Alarcon, J. M. Persistent modifications of hippocampal synaptic function during remote spatial memory. *Neurobiol. Learn. Mem.* **138**, 182–197 (2017).

12. 12.

Cimadevilla, J. M., Wesierska, M., Fenton, A. A. & Bures, J. Inactivating one hippocampus impairs avoidance of a stable room-defined place during dissociation of arena cues from room cues by rotation of the arena. *Proc. Natl Acad. Sci. USA* **98**, 3531–3536 (2001).

13. 13.

Hsieh, C. et al. Persistent increases of PKMzeta in memory-activated neurons trace LTP maintenance during spatial long-term memory storage. *Eur. J. Neurosci.* <https://doi.org/10.1111/ejn.15137> (2021).

14. 14.

Pastalkova, E. et al. Storage of spatial information by the maintenance mechanism of LTP. *Science* **313**, 1141–1144 (2006).

15. 15.

Tsokas, P. et al. Compensation for PKMzeta in long-term potentiation and spatial long-term memory in mutant mice. *eLife* **5**, e14846, (2016).

16. 16.

Talbot, Z. N. et al. Normal CA1 place fields but disordinated network discharge in a Fmr1-null mouse model of fragile X syndrome. *Neuron* **97**, 684–697 (2018).

17. 17.

van Dijk, M. T. & Fenton, A. A. On how the dentate gyrus contributes to memory discrimination. *Neuron* **98**, 832–845 (2018).

18. 18.

Wesierska, M., Dockery, C. & Fenton, A. A. Beyond memory, navigation, and inhibition: behavioral evidence for hippocampus-dependent cognitive coordination in the rat. *J. Neurosci.* **25**, 2413–2419 (2005).

19. 19.

Dvorak, D., Chung, A., Park, E. H. & Fenton, A. A. Dentate spikes and external control of hippocampal function. *Cell Rep.* **36**, 109497 (2021).

20. 20.

Whishaw, I. Q. & Gorny, B. Path integration absent in scent-tracking fimbria-fornix rats: evidence for hippocampal involvement in “sense of direction” and “sense of distance” using self-movement cues. *J. Neurosci.* **19**, 4662–4673 (1999).

21. 21.

Bures, J. et al. Dissociation of exteroceptive and idiothetic orientation cues: effect on hippocampal place cells and place navigation. *Philos. Trans. R. Soc. Lond. B* **352**, 1515–1524 (1997).

22. 22.

Jezek, K. et al. Stress-induced out-of-context activation of memory. *PLoS Biol.* **8**, e1000570 (2010).

23. 23.

Garner, A. R. et al. Generation of a synthetic memory trace. *Science* **335**, 1513–1516 (2012).

24. 24.

Witter, M. P. The perforant path: projections from the entorhinal cortex to the dentate gyrus. *Prog. Brain Res.* **163**, 43–61 (2007).

25. 25.

O'Reilly, K. C., Perica, M. I. & Fenton, A. A. Synaptic plasticity/dysplasticity, process memory and item memory in rodent models of mental dysfunction. *Schizophr. Res.* **207**, 22–36 (2019).

26. 26.

Nitz, D. & McNaughton, B. Differential modulation of CA1 and dentate gyrus interneurons during exploration of novel environments. *J. Neurophysiol.* **91**, 863–872 (2004).

27. 27.

Madar, A. D., Ewell, L. A. & Jones, M. V. Temporal pattern separation in hippocampal neurons through multiplexed neural codes. *PLoS Comput. Biol.* **15**, e1006932 (2019).

28. 28.

Mercer, L. F., Jr., Remley, N. R. & Gilman, D. P. Effects of urethane on hippocampal unit activity in the rat. *Brain Res. Bull.* **3**, 567–570 (1978).

29. 29.

Ferrante, M., Migliore, M. & Ascoli, G. A. Feed-forward inhibition as a buffer of the neuronal input-output relation. *Proc. Natl Acad. Sci. USA* **106**, 18004–18009 (2009).

30. 30.

Vasuta, C. et al. Metaplastic regulation of CA1 Schaffer collateral pathway plasticity by Hebbian mGluR1a-mediated plasticity at excitatory synapses onto somatostatin-expressing interneurons. *eNeuro* **2**, ENEURO.0051-15.2015 (2015).

31. 31.

Donato, F., Rompani, S. B. & Caroni, P. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).

32. 32.

Letzkus, J. J., Wolff, S. B. & Luthi, A. Disinhibition, a circuit mechanism for associative learning and memory. *Neuron* **88**, 264–276 (2015).

33. 33.

Ling, D. S. et al. Protein kinase Mzeta is necessary and sufficient for LTP maintenance. *Nat. Neurosci.* **5**, 295–296 (2002).

34. 34.

Takeuchi, T., Duszkiewicz, A. J. & Morris, R. G. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos. Trans. R. Soc. Lond. B* **369**, 20130288 (2013).

35. 35.

Lee, J. W. & Jung, M. W. Separation or binding? Role of the dentate gyrus in hippocampal mnemonic processing. *Neurosci. Biobehav. Rev.* **75**, 183–194 (2017).

36. 36.

Lee, J. W., Kim, W. R., Sun, W. & Jung, M. W. Role of dentate gyrus in aligning internal spatial map to external landmark. *Learn. Mem.* **16**, 530–536 (2009).

37. 37.

Luna, V. M. et al. Adult-born hippocampal neurons bidirectionally modulate entorhinal inputs into the dentate gyrus. *Science* **364**, 578–583 (2019).

38. 38.

Lee, H. et al. Early cognitive experience prevents adult deficits in a neurodevelopmental schizophrenia model. *Neuron* **75**, 714–724 (2012).

39. 39.

Aharoni, D. & Hoogland, T. M. Circuit investigations with open-source miniaturized microscopes: past, present and future. *Front. Cell Neurosci.* **13**, 141 (2019).

40. 40.

Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83–94 (2017).

41. 41.

Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).

42. 42.

Zhou, P. et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018).

43. 43.

Huang, L. et al. Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife* **10**, e51675 (2021).

44. 44.

Olypher, A. V., Lansky, P., Muller, R. U. & Fenton, A. A. Quantifying location-specific information in the discharge of rat hippocampal place cells. *J. Neurosci. Methods* **127**, 123–135 (2003).

45. 45.

Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 2nd edn (Academic Press, 2001).

46. 46.

Harrison, R. R. et al. A Low-power integrated circuit for a wireless 100-electrode neural recording system. *IEEE J. Solid State Circuits* **42**, 123–133 (2007).

47. 47.

Hrabetova, S. Extracellular diffusion is fast and isotropic in the stratum radiatum of hippocampal CA1 region in rat brain slices. *Hippocampus* **15**, 441–450 (2005).

48. 48.

Holsheimer, J. Electrical conductivity of the hippocampal CA1 layers and application to current-source-density analysis. *Exp. Brain Res.* **67**, 402–410 (1987).

49. 49.

Pettersen, K. H., Devor, A., Ulbert, I., Dale, A. M. & Einevoll, G. T. Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and

conductivity discontinuities. *J. Neurosci. Methods* **154**, 116–133 (2006).

50. 50.

Penttonen, M., Kamondi, A., Sik, A., Acsady, L. & Buzsaki, G. Feed-forward and feed-back activation of the dentate gyrus in vivo during dentate spikes and sharp wave bursts. *Hippocampus* **7**, 437–450 (1997).

51. 51.

Bragin, A., Jando, G., Nadasdy, Z., van Landeghem, M. & Buzsaki, G. Dentate EEG spikes and associated interneuronal population bursts in the hippocampal hilar region of the rat. *J. Neurophysiol.* **73**, 1691–1705 (1995).

52. 52.

Kanter, B. R. et al. A novel mechanism for the grid-to-place cell transformation revealed by transgenic depolarization of medial entorhinal cortex layer II. *Neuron* **93**, 1480–1492.e1486 (2017).

## Acknowledgements

Supported by NIH grants R01MH115304, R01NS105472, and R01AG043688. We thank T. Sacktor for valuable discussions.

## Author information

### Affiliations

1. Center for Neural Science, New York University, New York City, USA

Ain Chung, Alejandro Grau-Perales, Eliott R. J. Levy, Dino Dvorak, Nida Hussain & André A. Fenton

2. Department of Psychology, Hunter College, City University of New York, New York City, USA

Claudia Jou

3. Neuroscience Institute at the NYU Langone Medical Center, New York City, USA

André A. Fenton

## Contributions

A.A.F. and A.C. designed experiments. E.L. and A.G.-P. performed the calcium imaging experiments and analyses. A.C., C.J., D.D. and A.G.-P. performed behavioural experiments and analyses. A.C. and D.D. performed in vivo physiology experiments and analyses; A.C. and C.J. performed ex vivo physiology experiments and analyses; A.C. and A.G.-P. performed immunohistochemistry experiments. N.H. contributed to behaviour and preparation for in vivo physiology experiments. A.A.F., A.C., C.J., E.L. and D.D. analysed and interpreted the results. A.A.F. and A.C. wrote the manuscript with contributions from C.J., D.D., and E.L. A.A.F. directed and obtained research funding.

## Corresponding author

Correspondence to [André A. Fenton](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Active place avoidance and control task behaviors.

**a)** Task variants: The apparatus and environment were essentially the same for the three task variants. More than 99% of the time in the apparatus, when there was no shock, the visual environment was identical across the tasks; only the place learning environment was different in that the arena surface was covered in shallow water. During the 500-ms shock the cognitive control training and place learning groups experienced an unpleasant foot shock (depicted as red shaded sectors), whereas the spatial exploration group did not. **b)** Example paths during select trials of the same representative mice; dots indicate the mouse's location when it was shocked (red), and when it would have been shocked if the shock was on (black). **c)** The CCT (n=12), PL (n=11), and SE (N=9) groups differ in the initial but not the subsequent pretraining trials with no shock, because mice walk less in the shallow water (Group x Trial two-way RM ANOVA; group:  $F_{2,27} = 27.7$ ,  $p = 10^{-7}$ ;  $\eta^2 = 0.49$ ; Trial:  $F_{1,50.5} = 2.44$ ,  $p = 0.12$ ; interaction:  $F_{2,50.5} = 5.88$ ,  $p = 0.005$ ,  $\eta^2 = 0.19$ ; PL during initial < all other group x trials measures). This resulted in group differences in the number of times the mice enter the location of the future shock zone (Group x Trial two-way RM ANOVA, Group:  $F_{2,31} = 10.51$ ,  $p = 10^{-4}$ ,  $\eta^2 = 0.22$ ; Trial:  $F_{1,60.7} = 0.042$ ,  $p = 0.8$ ; Group X Trial:  $F_{2,60.7} = 2.93$ ,  $p = 0.06$ ; Tukey-HSD: \*\*\*CCT = SE > PL Consequently, performance in the initial and subsequent training sessions is analyzed separately. The group average performance measured during 30-min by entrances into the shock zone or equivalent area for the SE mice. Statistical analysis of subsequent training (see Supplemental Information for analysis of initial training; \* $p < 0.05$ ):

Group X Trials 1-3 two-way RM ANOVA, Group:  $F_{2,30} = 0.04$ ,  $p = 0.97$ ; Trials:  $F_{2,56.8} = 12.3$ ,  $p = 10^{-5}$ ,  $\eta^2 = 0.19$ , Tukey-HSD:  $Ts1 > Ts2 = Ts3$ ; Interaction:  $F_{4,63.2} = 0.93$ ,  $p = 0.46$ ). Place avoidance learning measured as decreasing entrances into the shock zone was robust with significant effects of Group and the Group X Trial interaction (Group:  $F_{2,45.6} = 72.3$ ,  $p = 10^{-11}$ ,  $\eta^2 = 0.23$ ; Trial:  $F_{2,34.7} = 4.63$ ,  $p = 0.02$ ,  $\eta^2 = 0.006$ ; Interaction:  $F_{4,39.3} = 8.31$ ,  $p = 10^{-5}$ ,  $\eta^2 = 0.01$ ; \*\*\*Tukey-HSD: SE > CCT = PL). The two conditioned groups showed similar learning (Group:  $F_{1,59.2} = 0.93$ ,  $p = 0.34$ ,  $\eta^2 = 0.004$ ; Trial:  $F_{2,33.5} = 39.5$ ,  $p = 10^{-9}$ ,  $\eta^2 = 0.20$ , Tukey-HSD Trial1 > Trial 2 > Trial 3; Interaction:  $F_{2,33.5} = 1.57$ ,  $p = 0.2$ ,  $\eta^2 = 0.01$ ). The groups differed on the 1-week retention test ( $F_{2,28} = 3.57$ ,  $p = 0.04$ ,  $\eta^2 = 0.21$ ). Tukey-HSD comparisons of groups: \* $p < 0.05$ . d) Total path walked measures activity: During initial training the conditioned mice walk less and restrict themselves to the arena periphery compared to the SE mice; the conditioned mice did not differ in the distance they walked (Group x Trial two-way RM ANOVA, Group:  $F_{2,41.2} = 11.51$ ,  $p = 10^{-4}$ ,  $\eta^2 = 0.14$ ; Trial:  $F_{2,28.5} = 1.16$ ,  $p = 0.33$ ,  $\eta^2 = 0.002$ ; Group X Trial:  $F_{4,32.4} = 5.83$ ,  $p = 0.001$ ,  $\eta^2 = 0.03$ ; SE > CCT = PL). e) Time to 1<sup>st</sup> Entrance estimates between-session memory. One-week memory retention was robust with an obvious effect of Group measured by entrances into the shock zone (see Fig. 2b;  $F_{2,31} = 101.3$ ,  $p = 10^{-14}$ ,  $\eta^2 = 0.73$ , \*\*\*Tukey-HSD: SE > CCT = PL) as well as the time to first enter the shock zone in the two conditioned groups with PL better than CCT (Group:  $F_{1,31} = 3.25$ , \* $p = 0.05$ ,  $\eta^2 = 0.21$ ). Line plot: mean±s.e.m.

### [Source data](#)

## **Extended Data Fig. 2 Active place avoidance in a novel environment is improved by prior CCT training.**

a left) Times to enter the shock zone for the first five times during subsequent training in entirely novel physical conditions (Ts1) are prolonged after initial training (Ti1) in the CCT group (n=12) comparison

of slopes:  $F_{1,116} = 15.91$ ,  $p = 0.0001$ ,  $\eta^2 = 0.12$ ). a middle) CCT mice also prolonged entering the shock zone during subsequent training more than the other groups (comparison of slopes:  $F_{2,154} = 5.94$ ,  $p = 0.003$ ,  $\eta^2 = 0.07$ ). a right) PL ( $n=11$ ) and SE ( $n=9$ ) groups performed indistinguishably during their Ts1 from the CCT Ti1 (comparison of slopes:  $F_{2,154} = 2.60$ ,  $p = 0.07$ ). Solid lines depict linear regressions, and dotted lines depict 95% confidence interval bounds. **b)** The number of shock zone entrances during the first 5 min of CCT training. CCT Ti1 vs Ts1: paired  $t_{11} = 2.58$ , 1-tailed \* $p = 0.01$ ,  $d = 0.75$ . CCT Ts1 vs PL Ts1:  $t_{21} = 2.48$ , 1-tailed \* $p = 0.01$ ,  $d = 1.04$ ; CCT Ts1 vs SE:  $t_{19} = 2.46$ , 1-tailed \* $p = 0.01$ ,  $d = 1.09$ ), which are all significant after Bonferroni corrections for the multiple (3) tests. CCT ( $n=12$ ), PL ( $n=11$ ), and SE ( $N=9$ ). Box plot: 25-75%, median, whiskers: min/max. **c)** Schematic training protocol. **d)** Learned place avoidance during initial training predicts subsequent place avoidance in a novel environment: If place avoidance during initial training facilitates learning during subsequent training, then mice that express better learned place avoidance in initial training (measured as performance in RET) are predicted to show better performance during the early stages of subsequent training (Ts1). One-week retention of the conditioned place avoidance significantly predicts place avoidance in a novel environment amongst the CCT (left,  $n=12$ ) but not the PL (right,  $n=11$ ) mice. **e)** Similarly, if initial CCT facilitates subsequent learning in the novel environment, this predicts that if mice perform well during the early stages of subsequent training (Ts1) then they would have learned well as estimated by performance at the end of initial training (Ti3). Correlation between initial training trials (Ti1-3) and during subsequent (Ts1) training increased from Ti1 to Ti3 in the CCT group, but not within the PL group. Drawing Source: Ain Chung, André Fenton.

### Source data

### Extended Data Fig. 3 Targeting electrodes to investigate pathway-specific hippocampal responses to medial perforant path simulation.

**a–c)** Muscimol inactivation demonstrates DG is crucial for expression of active place avoidance several days after CCT training: **a)** Experimental

design (1) and the tracks of an exemplar muscimol-injected mouse during each behavioral session (2). **b**) Schematic from Franklin and Paxinos<sup>45</sup> illustrating the injection and guide cannulae target, a histology section of the cannulation track, as well as a histological section after injecting fluorogold (FG) to estimate the infusion spread (scale bar: 500μm). **c**) A measure of memory expression illustrates that targeting muscimol at DG reversibly impairs established active place avoidance memory (n = 5/group, Group X Retention session RM ANOVA Group:  $F_{1,8} = 2.46$ , p = 0.16; Retention session:  $F_{2,16} = 5.84$ , p = 0.01,  $\eta^2 = 0.42$ ; interaction:  $F_{2,16} = 4.01$ , p = 0.03,  $\eta^2 = 0.33$ , two-sided Šidak's multiple comparisons RET2 < RET1 = RET3 in Musimol injected group, Musimol vs PBS in RET2, \*\* p <0.01. Box plot: 25-75%, median, whiskers: min/max. **d**) Targeting of the recording and stimulating electrodes: Schematic and Nissl-stained histological sections from an exemplar mouse illustrating the left) recording and right) stimulation electrode sites. (Scale bars: 1mm) Schematics from Franklin and Paxinos<sup>45</sup>. **e**) Membrane potential vs. current source density (CSD) plots to illustrate signal localization: comparison of the raw field potential (mV) plots and corresponding CSD plots along corresponding channels of the linear silicon probe. Passive volume conduction strongly influences the signal recorded at adjacent electrodes in the field potential plots. Some examples are indicated by the red/blue boxes. 1: signal artifacts in CA1; 2: signal artifact at the hippocampal fissure; 3: volume conduction from MPP responses that occlude LPP responses; 4: volume conduction from the granule cell layer response occluding hilar responses. **f**) Representative sink and source signals and their locations from the CSD plots. **g–j**) Confirmation that electrical stimulation targeted the MPP. **g**) Nissl section and anti-human antigen (HA) immunostaining of transgenic mice (TRE-hM4Di-tetO+/- crossed with Ent-tTA+/-) that express the inhibitory DREADD under control of doxycycline withdrawal. The inhibitory DREADD immunoreactivity localizes to hippocampus-projecting MECII neurons. Immunofluorescence shows expression in MECII cell bodies, in the perforant path axons and in their termination zones in the molecular layer of the dentate gyrus and CA3. Expression is faint in CA1 *str. lacunosum moleculare*. Expression is in ~20% of MECII stellate cells and 5% in MECIII cells with less in adjacent regions (Kanter et al. 2017). **h**) Experimental design to evaluate effective stimulation of the MPP. **i**)

Average evoked responses recorded with a 16-site linear silicon probe spanning the somato-dendritic axis of dorsal hippocampus. **j**) Input-output curves of fEPSP slope responses recorded from the DG molecular layer and population spike responses recorded from the DG hilus. These data quantify that responses to electrical stimulation targeted to the angular bundle is dose-dependently and reversibly suppressed by CNO-mediated activation of hM4Di in the MPP (Area X Dose RM ANOVA, Area:  $F_{1,8} = 7.55$ ,  $p = 0.02$ ,  $\eta^2 = 0.49$ ; Dose:  $F_{4,32} = 168.6$   $p = 10^{-21}$ ,  $\eta^2 = 0.95$ ; interaction:  $F_{4,32} = 5.78$ ,  $p = 0.001$ ,  $\eta^2 = 0.42$ , two-sided Šidák-corrected within-area comparisons \*\*\* $p < 0.001$ ). Box plot: 25-75%, median, whiskers: min/max. Line plot: mean±s.e.m. Drawing Source: Ain Chung, André Fenton.

### Source data

### Extended Data Fig. 4 Additional active place avoidance training does not incrementally change the response to medial perforant path stimulation.

**a)** Experimental design, **b)** Learning curves comparing initial CCT training, and subsequent CCT training to either a conflicting location of shock (Conflict CCT), or to a novel location of shock in a novel room (subsequent CCT). (Number of Entrances: Group X Trial RM ANOVA Group:  $F_{2,34} = 2.95$ ;  $p = 0.07$ ; Trial:  $F_{3,99} = 22.56$ ,  $p = 10^{-5}$ ,  $\eta^2 = 0.41$ ; Interaction;  $F_{6,99} = 5.98$ ,  $p = 10^{-5}$ ,  $\eta^2 = 0.27$ , Time to 1<sup>st</sup> Entrance: Group X Trial RM ANOVA Group: :  $F_{2,34} = 4.42$ ,  $p = 0.02$ ,  $\eta^2 = 0.21$  Trial:  $F_{3,99} = 13.57$ ;  $p = 10^{-5}$ ,  $\eta^2 = 0.29$ ; Interaction;  $F_{6,99} = 1.581$ ,  $p = 0.16$ ). (initial and subsequent CCT n=12, conflict group n=11). Line plot: mean±SEM. **c)** CCT (n=7) caused changes at the corresponding middle molecular layers of the supraDG but not the infraDG response to MPP stimulation (Area X Training RM ANOVA, area:  $F_{1,12} = 0.49$ ,  $p = 0.50$ ; training:  $F_{1,12} = 11.08$ ,  $p = 0.006$ ,  $\eta^2 = 0.48$ ; interaction:  $F_{1,12} = 6.69$ ,  $p = 0.02$ ,  $\eta^2 = 0.36$ ; Šidák's multiple comparisons: only SupraDG changes circuit function \*\* $p = 0.003$ ; supraDG was significant paired  $t_6 = 5.09$ , \*\* $p = 0.002$ ,  $d = 1.92$  but not at infraDG sites (paired  $t_6 = 0.46$ ,  $p = 0.66$ ). Scale bar: 50  $\mu$ A/mm<sup>3</sup>, 5 ms. **d)** No further

changes are observed in the synaptic responses after either additional CCT in a third, novel environment ( $n=5$ , supraDG: paired  $t_4 = 0.45$ ,  $p = 0.67$ ; infraDG paired  $t_4 = 0.12$ ,  $p = 0.90$ ) or in the same environment with a 180°-relocated shock zone that conflicts with the prior location of shock ( $n=5$ , supraDG paired  $t_4 = 0.02$ ,  $p = 0.98$ ; infraDG paired  $t_4 = 1.13$ ,  $p = 0.32$ ). The 4 mice/group that received both types of subsequent training, received conflict or novel environment training, in counterbalanced order. Box plot: 25-75%, median, whiskers: min/max. Line plot: mean±s.e.m. Drawing Source: Ain Chung, André Fenton.

#### Source data

#### Extended Data Fig. 5 Stability of electrode properties and placement, and time course of the learning-induced changes in the response to medial perforant path stimulation.

**a–c)** Stability of stimulation and recording electrode properties across the duration of the experiment. **a)** Average power of spontaneous theta oscillations measured at the molecular-hilar region of supraDG and **b)** stimulation electrode impedance, both measured during wheel running before active place avoidance training (see schematic at top). Neither theta power ( $n=8$ ,  $F_{2,4,21.6} = 0.65$ ,  $p = 0.55$ ) nor electrode impedance in 32 channels ( $F_{4,128} = 0.294$ ,  $p = 0.9$ ) changed across the experiment, making it unlikely that changed electrode properties account for the CCT-induced differences in evoked responses. Box plot: 25-75%, median, whiskers: min/max. **c)** Session-averaged current source density (CSD) of LPP-triggered DS<sub>L</sub> (type 1) and MPP-triggered DS<sub>M</sub> (type2) spontaneous events also illustrate recording mechanical stability across the experiment. **d)** Time course of CCT-induced changes in MPP-DG synaptic responses. fEPSP slope was normalized to the baseline slope prior to exposure to the behavioral arena. The decrease of the fEPSP response is observed within 2 h of the first CCT trial. \* indicates a significant difference from before to after a CCT trial ( $n = 7$ , Paired t-test:  $t_6 = 3.40$ , \* $p = 0.01$ ,  $d = 1.28$ ) Box plot: 25-75%, median, whiskers: min/max. **e)** CCT causes place avoidance and reduced population synaptic responses to MPP stimulation that persist

at least 60 days. Left) Conditioned place avoidance ( $n = 3$ ,  $F_{1,109,2,038} = 31.07$ ,  $p = 0.03$ ,  $\eta^2 = 0.94$ , pretraining > 1-wk retention = 60-d retention) and right) population synaptic responses ( $F_{1,015,2,030} = 19.30$ ,  $p = 0.05$ ,  $\eta^2 = 0.90$ , baseline > 1-wk = 60-d, Tukey-HSD \* $p<0.05$ ). Drawing Source: Ain Chung, André Fenton.

[Source data](#)

**Extended Data Fig. 6 Naturally occurring, spontaneous dentate spikes originating from MPP activation confirm the findings using evoked responses to MPP stimulation.**

**a)** A dentate spike (DS, stars) is the DG response to synchronous input from the medial and lateral perforant path, making it a physiological estimate of the MPP-DG connection that is independent of artificial stimulation. DS were identified in the DG hilus as sharp, positive waves of the local field potential with prominence (distance between peak and closest preceding or following trough) greater than 2 S.D. of all detected positive peaks, as well as width between 2.5 and 12.5 ms measured at 50% of the peak's prominence. **b)** Two types of dentate spikes –  $DS_L$  (type 1) and  $DS_M$  (type 2) were identified using a CSD fingerprinting method, where all DS that exhibited a symmetric pair of current sinks in the distal molecular layers of DG were identified as  $DS_L$  and DS that exhibited a symmetric pair of current sinks in the medial/proximal molecular layers of DG were identified as  $DS_M$ . The average CSD profiles of  $DS_L$  (left) and  $DS_M$  (right) computed from all classified DS events with clearly distinct pairs of current sinks in the distal and medial/proximal molecular layers of DG respectively. **c)** Average CSD profiles of DS-associated current sinks in molecular layers of DG in supraDG (top row) and infraDG (bottom row) triggered by  $DS_L$  (left column) and  $DS_M$  (right column). Black and red colors represent before and after CCT training, respectively. **c)** Summary comparisons of maximum gradient of the DS sink (i.e. negative slope) measured before and after CCT. CCT only changed  $DS_M$  (type 2) at the supraDG site but not at the infraDG site, confirming findings assessed by stimulus evoked responses ( $n=5$ ,

Paired t-test:  $t_4 = 3.04$ , \* $p = 0.04$ ,  $d = 1.36$ ). Box plot: 25-75%, median, whiskers: min/max.

Source data

**Extended Data Fig. 7 Characterization, hypotheses, and validation of experiments to assess the mechanism of the attenuated response to MPP stimulation after CCT.**

**a)** Gad2-expressing interneurons are more abundant in supraDG than infraDG. Left) eYFP expressing interneurons in Gad2Cre-eYFP mice (scale bar 0.2 mm); Right) cell counts per section in mice after CCT ( $n=6$ ) or SE ( $n=4$ ) experience document enrichment of inhibitory interneurons in supraDG relative to infraDG ( $F_{1.215,10.93} = 21.40$ ,  $p < 0.001$ ,  $\eta^2 = 0.69$ , SupraDG > InfraDG = Hilus, Tukey-HSD: \*\*\* $p = 0.0008$ , \*\* $p = 0.003$ ). Orange and green arrows indicate eYFP+ cells in the supraDG and infraDG, respectively. Box plot: 25-75%, median, whiskers: min/max. **b)** Two hypotheses to account for the CCT-induced suppression of DG responses to MPP stimulation: Schematic of MPP→granule cell (GC) depression and potentiation of GC inhibition hypotheses to explain the CCT-induced reduction of MPP response. **c–i)** The *in vivo* experiment under urethane anesthesia to investigate the role of inhibition in the CCT-induced effects. **c)** Experimental design, **d)** schematic of the preparation, **e)** photo illustrating the recording electrodes and optical stimulation, and **f)** The schematic is superimposed on a histological section of a Gad2-Cre-eYFP mouse's dorsal hippocampus with eYFP (green) and immunofluorescence for DAPI (blue), to indicate the circuit that was targeted (scale bar 0.02mm). **e)** Atlas from Franklin and Paxinos<sup>45</sup> and **g)** corresponding histological sections illustrating electrode and optical fiber placements (scale bar:1mm). Sixteen-channel linear electrode array recordings of LFPs across the somatodendritic axis of dorsal hippocampus **h)** under urethane anesthesia and **i)** in the freely-behaving mouse. Under urethane, LFPs reflecting rhythmic ongoing synaptic activity is much attenuated compared to in the behaving mouse. Voltage scale (red bar) = 1 mV.

[Source data](#)

**Extended Data Fig. 8 validation of optogenetic activation of inhibition to evaluate whether altered inhibition contributes to the CCT-induced changes of the response to MPP stimulation.**

**a)** Population spike responses are blocked by activating ChR2 in Gad2-expressing cells 5 ms before MPP stimulation *in vivo* under urethane anesthesia, regardless of prior training. Input-output relationships of the response to MPP stimulation with and without prior light activation of inhibitory-neuron ChR2 in the (upper) supraDG and (lower) infraDG. This demonstrates light-stimulated inhibition is effective in all groups. home cage (n = 5), SE (n = 6), PL (n = 4), and CCT mice (n = 7) **b)** fEPSP evoked responses at supraDG are reduced by activating Gad2-expressing cells at the same time as MPP stimulation *in vivo*, regardless of prior training. Upper) Input-output relationships of the response to MPP stimulation with and without concurrent light activation of inhibitory-neuron ChR2. Lower) The corresponding summaries represented as the area-under-the-curve (AUC) for statistical analysis of the group and light stimulation effects. Group X Stimulation ANOVA, Group:  $F_{3,13} = 0.80$ , p = 0.51; Stimulation:  $F_{1,13} = 16.05$ , p = 0.0015,  $\eta^2 = 0.55$ ; interaction:  $F_{2,10} = 0.4921$ , p = 0.70). home cage (n = 5), SE (n = 4), PL (n = 4), and CCT mice (n = 4) Box plot: 25-75%, median, whiskers: min/max. Line plot: mean±s.e.m.

[Source data](#)

**Extended Data Fig. 9 *in vivo* paired-pulse inhibition is weaker after CCT.**

**a)** CSD examples of single responses to paired-pulse inhibition in CCT, homecage, SE and PL mice (Scale bars: 100  $\mu$ A/mm<sup>3</sup>, 5 ms); MPP-stimulated inhibition is attenuated at supraDG of CCT mice. **b)** Summary PPI curves and indices (supraDG:  $F_{3,18} = 3.22$ , \*p = 0.04,  $\eta^2 = 0.15$ ; infraDG:  $F_{3,18} = 0.95$ , p = 0.43). **b)** Responses to MPP stimulation were

recorded in urethane-anesthetized mice at the supra and infra granule cell layers of the DG. Significant changes were not detected although there is a hint of enhanced facilitation in the CCT group. (supraDG granule cell Group:  $F_{3,18} = 0.47$ ,  $p = 0.70$ ; InfraDG granule cell group:  $F_{3,18} = 0.19$ ,  $p = 0.90$ ). home cage ( $n = 5$ ), SE ( $n = 6$ ), PL ( $n = 4$ ), and CCT mice ( $n = 7$ ) Box plot: 25-75%, median, whiskers: min/max. Line plot: mean $\pm$ s.e.m.

### [Source data](#)

### [Extended Data Fig. 10 Validation of MPP-stimulation related optogenetic terminal activation of inhibition, and experience-related changes of MPP-stimulated responses in spontaneously-inactive \*ex vivo\* hippocampus slices.](#)

**a)** Experimental design and photomicrograph of *ex vivo* hippocampus slice prepared 1-week posttraining (scale bar: 200 $\mu$ m). fEPSP responses to MPP terminal stimulation recorded with light-activated ChR2 in Gad2+ cells. **b)** Response recorded in infraDG to MPP stimulation with ChR2 activation of inhibitory neurons at different time offsets. left: infraDG (2-way RM ANOVA Group:  $F_{2,17} = 0.50$ ,  $p = 0.62$ , Light-Stim ISI  $F_{2,06,34.95} = 2.82$ ,  $p = 0.73$ ; interaction:  $F_{10,85} = 1.15$ ,  $p = 0.34$ ) right: Light 2ms before ( $F_{2,17} = 1.14$ ,  $p = 0.34$ ). SE ( $n = 6$ ), PL ( $n = 8$ ), and CCT mice ( $n = 6$ ) **c)** Input-output relationship of the response to MPP stimulation in *ex vivo* hippocampal slices. I-O curves (left) and their summary as the area-under-the-curve (AUC, right) upon which statistical group comparisons were made. (Upper) SupraDG: IO-AUC  $F_{3,32} = 3.80$ ,  $p = 0.02$ ,  $\eta^2 = 0.26$ , Šidák's multiple comparisons: HC vs PL \* $p = 0.021$ . (Lower) InfraDG: IO-AUC  $F_{3,31} = 0.93$ ,  $p = 0.44$  homecage ( $n = 6$ ) SE ( $n = 10$ ), PL (SupraDG  $n = 8$ , infraDG  $n = 9$ ), and CCT mice ( $n = 10$ ) Box plot: 25-75%, median, whiskers: min/max. Line plot: mean $\pm$ s.e.m.

### [Source data](#)

## Supplementary information

## **Reporting Summary**

## **Source data**

**Source Data Fig. 1**

**Source Data Fig. 2**

**Source Data Fig. 3**

**Source Data Fig. 4**

**Source Data Extended Data Fig. 1**

**Source Data Extended Data Fig. 2**

**Source Data Extended Data Fig. 3**

**Source Data Extended Data Fig. 4**

**Source Data Extended Data Fig. 5**

**Source Data Extended Data Fig. 6**

**Source Data Extended Data Fig. 7**

**Source Data Extended Data Fig. 8**

**Source Data Extended Data Fig. 9**

**Source Data Extended Data Fig. 10**

## **Rights and permissions**

## [Reprints and Permissions](#)

## About this article

### Cite this article

Chung, A., Jou, C., Grau-Perales, A. *et al.* Cognitive control persistently enhances hippocampal information processing. *Nature* **600**, 484–488 (2021). <https://doi.org/10.1038/s41586-021-04070-5>

- Received: 13 September 2019
- Accepted: 29 September 2021
- Published: 10 November 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04070-5>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04070-5>

- Article
- [Published: 24 November 2021](#)

# Contextual inference underlies the learning of sensorimotor repertoires

- [James B. Heald](#) ORCID: orcid.org/0000-0002-7293-7914<sup>1,2,3</sup>,
- [Máté Lengyel](#)<sup>2,4</sup> &
- [Daniel M. Wolpert](#) ORCID: orcid.org/0000-0001-7266-0049<sup>1,2,3</sup>

*Nature* volume 600, pages 489–493 (2021)

- 5407 Accesses
- 1 Citations
- 119 Altmetric
- [Metrics details](#)

## Subjects

- [Human behaviour](#)
- [Motor control](#)

## Abstract

Humans spend a lifetime learning, storing and refining a repertoire of motor memories. For example, through experience, we become proficient at manipulating a large range of objects with distinct dynamical properties.

However, it is unknown what principle underlies how our continuous stream of sensorimotor experience is segmented into separate memories and how we adapt and use this growing repertoire. Here we develop a theory of motor learning based on the key principle that memory creation, updating and expression are all controlled by a single computation—contextual inference. Our theory reveals that adaptation can arise both by creating and updating memories (proper learning) and by changing how existing memories are differentially expressed (apparent learning). This insight enables us to account for key features of motor learning that had no unified explanation: spontaneous recovery<sup>1</sup>, savings<sup>2</sup>, anterograde interference<sup>3</sup>, how environmental consistency affects learning rate<sup>4,5</sup> and the distinction between explicit and implicit learning<sup>6</sup>. Critically, our theory also predicts new phenomena—evoked recovery and context-dependent single-trial learning—which we confirm experimentally. These results suggest that contextual inference, rather than classical single-context mechanisms<sup>1,4,7,8,9</sup>, is the key principle underlying how a diverse set of experiences is reflected in our motor behaviour.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

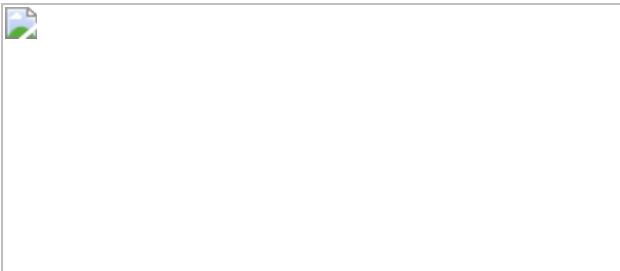
[Rent or Buy](#)

All prices are NET prices.

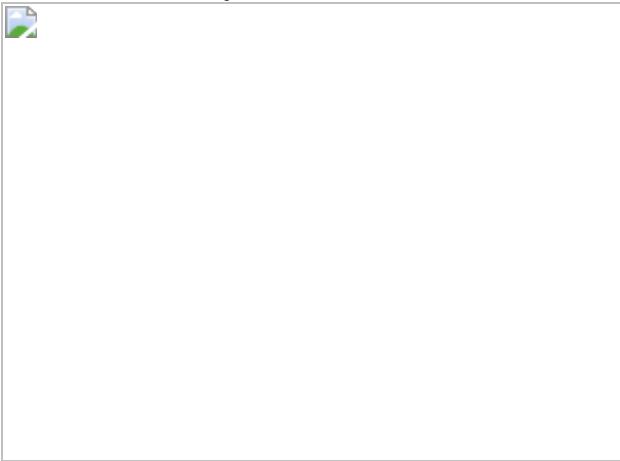
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

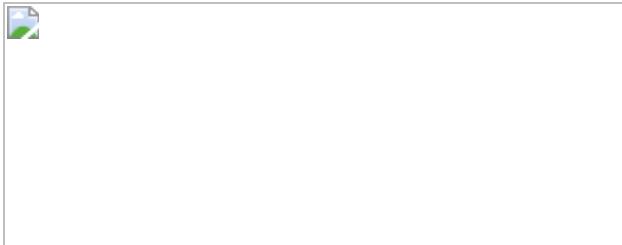
**Fig. 1: Contributions of contextual inference to motor learning in the COIN model.**



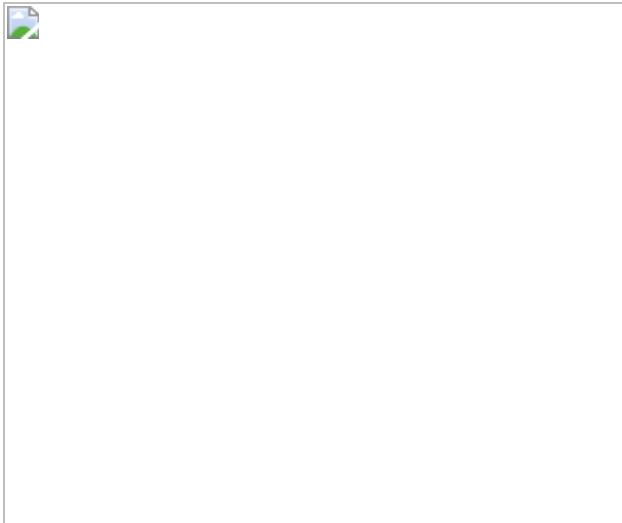
**Fig. 2: Memory creation and expression accounts for spontaneous and evoked recovery.**



**Fig. 3: Memory updating depends on contextual inference.**



**Fig. 4: Contextual inference underlies apparent changes in learning rate.**



## Data availability

All experimental data are publicly available at the Dryad repository (<https://doi.org/10.5061/dryad.m63xsj42r>). The data include the raw kinematics and force profiles of individual participants on all trials as well as the adaptation measures used to generate the experimental data shown in Figs. 2c, e and 3d.

## Code availability

The code for the COIN model is available at GitHub (<https://github.com/jamesheald/COIN>).

## References

1. 1.

Smith, M. A., Ghazizadeh, A. & Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* **4**, e179 (2006).

2. 2.

Kitago, T., Ryan, S., Mazzoni, P., Krakauer, J. W. & Haith, A. M. Unlearning versus savings in visuomotor adaptation: comparing effects of washout, passage of time, and removal of errors on motor memory. *Front. Hum. Neurosci.* **7**, 307 (2013).

3. 3.

Sing, G. C. & Smith, M. A. Reduction in learning rates associated with anterograde interference results from interactions between different timescales in motor adaptation. *PLoS Comput. Biol.* **6**, e1000893 (2010).

4. 4.

Herzfeld, D. J., Vaswani, P. A., Marko, M. K. & Shadmehr, R. A memory of errors in sensorimotor learning. *Science* **345**, 1349–1353 (2014).

5. 5.

Gonzalez Castro, L. N., Hadjiosif, A. M., Hemphill, M. A. & Smith, M. A. Environmental consistency determines the rate of motor adaptation. *Curr. Biol.* **24**, 1050–1061 (2014).

6. 6.

McDougle, S. D. et al. Credit assignment in movement-dependent reinforcement learning. *Proc. Natl Acad. Sci. USA* **113**, 6797–6802 (2016).

7. 7.

Donchin, O., Francis, J. T. & Shadmehr, R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J. Neurosci.* **23**, 9032–9045 (2003).

8. 8.

Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).

9. 9.

Shadmehr, R., Smith, M. A. & Krakauer, J. W. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108 (2010).

10. 10.

Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).

11. 11.

Oh, Y. & Schweighofer, N. Minimizing precision-weighted sensory prediction errors via memory formation and switching in motor adaptation. *J. Neurosci.* **39**, 9237–9250 (2019).

12. 12.

Gershman, S. J., Radulescu, A., Norman, K. A. & Niv, Y. Statistical computations underlying the dynamics of memory updating. *PLoS Comput. Biol.* **10**, e1003939 (2014).

13. 13.

Hulst, T. et al. Cerebellar degeneration reduces memory resilience after extended training. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.07.03.185959> (2021).

14. 14.

Pekny, S. E., Criscimagna-Hemminger, S. E. & Shadmehr, R. Protection and expression of human motor memories. *J. Neurosci.* **31**, 13829–13839 (2011).

15. 15.

Rescorla, R. A. & Heth, C. D. Reinstatement of fear to an extinguished conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process.* **1**, 88–96 (1975).

16. 16.

Berniker, M. & Kording, K. Estimating the sources of motor errors for adaptation and generalization. *Nat. Neurosci.* **11**, 1454–1461 (2008).

17. 17.

Taylor, J. A., Wojaczynski, G. J. & Ivry, R. B. Trial-by-trial analysis of intermanual transfer during visuomotor adaptation. *J. Neurophysiol.* **106**, 3157–3172 (2011).

18. 18.

Kording, K. P., Tenenbaum, J. B. & Shadmehr, R. The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat. Neurosci.* **10**, 779–786 (2007).

19. 19.

Heald, J. B., Ingram, J. N., Flanagan, J. R. & Wolpert, D. M. Multiple motor memories are learned to control different points on a tool. *Nat. Hum. Behav.* **2**, 300–311 (2018).

20. 20.

Coltman, S. K., Cashaback, J. G. A. & Gribble, P. L. Both fast and slow learning processes contribute to savings following sensorimotor

adaptation. *J. Neurophysiol.* **121**, 1575–1583 (2019).

21. 21.

Huang, V. S., Haith, A., Mazzoni, P. & Krakauer, J. W. Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron* **70**, 787–801 (2011).

22. 22.

Keisler, A. & Shadmehr, R. A shared resource between declarative memory and motor memory. *J. Neurosci.* **30**, 14817–14823 (2010).

23. 23.

McDougle, S. D., Ivry, R. B. & Taylor, J. A. Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends Cogn. Sci.* **20**, 535–544 (2016).

24. 24.

McDougle, S. D., Bond, K. M. & Taylor, J. A. Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *J. Neurosci.* **35**, 9568–9579 (2015).

25. 25.

Miyamoto, Y. R., Wang, S. & Smith, M. A. Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nat. Neurosci.* **23**, 443–455 (2020).

26. 26.

Haruno, M., Wolpert, D. M. & Kawato, M. MOSAIC model for sensorimotor learning and control. *Neural Comput.* **13**, 2201–2220 (2001).

27. 27.

Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).

28. 28.

Sanders, H., Wilson, M. A. & Gershman, S. J. Hippocampal remapping as hidden state inference. *eLife* **9**, e51140 (2020).

29. 29.

Collins, A. & Koechlin, E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* **10**, e1001293 (2012).

30. 30.

Collins, A. G. E. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).

31. 31.

Howard, I. S., Ingram, J. N. & Wolpert, D. M. A modular planar robotic manipulandum with end-point torque control. *J. Neurosci. Methods* **181**, 199–211 (2009).

32. 32.

Milner, T. E. & Franklin, D. W. Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *J. Physiol.* **567**, 651–664 (2005).

33. 33.

Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Rymer, W. Z. & Mussa-Ivaldi, F. A. Persistence of motor adaptation during constrained, multi-joint, arm movements. *J. Neurophysiol.* **84**, 853–862 (2000).

34. 34.

Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. Hierarchical Dirichlet processes. *J. Am. Stat. Assoc.* **101**, 1566–1581 (2006).

35. 35.

Fox, E. B., Sudderth, E. B., Jordan, M. I. & Willsky, A. S. An HDP-HMM for systems with state persistence. In *Proc. 25th International Conference on Machine Learning* (eds. McCallum, A. & Roweis, S.) 312–319 (Omnipress, 2008).

36. 36.

Teh, Y. W. in *Encyclopedia of Machine Learning* (eds. Sammut, C. & Webb, G. I.) 280–287 (Springer, 2011).

37. 37.

Carvalho, C. M., Johannes, M. S., Lopes, H. F. & Polson, N. G. Particle learning and smoothing. *Stat. Sci.* **25**, 88–106 (2010).

38. 38.

Lopes, H. F., Carvalho, C. M., Johannes, M. S., & Polson, N. G. in *Bayesian Statistics 9* (eds. Bernardo, J. M. et al.) (Oxford Univ. Press, 2011).

39. 39.

Houlsby, N. et al. Cognitive tomography reveals complex, task-independent mental representations. *Curr. Biol.* **23**, 2169–2175 (2013).

40. 40.

Acerbi, L. & Ma, W. J. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. In *Proc. Advances in Neural Information Processing Systems* (eds. Guyon, I. et al.) 1836–1846 (Curran, 2017).

41. 41.

Jeffreys, H. *The Theory of Probability* (Oxford Univ. Press, 1998).

42. 42.

Li, J., Wang, Z. J., Palmer, S. J. & McKeown, M. J. Dynamic Bayesian network modeling of fMRI: a comparison of group-analysis methods. *Neuroimage* **41**, 398–407 (2008).

## Acknowledgements

We thank J. N. Ingram for technical support and G. Hennequin for discussions. This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 726090 to M.L.), the Wellcome Trust (Investigator Awards 212262/Z/18/Z to M.L. and 097803/Z/11/Z to D.M.W.), the Royal Society (Noreen Murray Professorship in Neurobiology to D.M.W.), the National Institutes of Health (R01NS117699 and U19NS104649 to D.M.W.) and Engineering and Physical Sciences Research Council (studentship to J.B.H.).

## Author information

### Author notes

1. These authors contributed equally: Máté Lengyel, Daniel M. Wolpert

## Affiliations

1. Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA

James B. Heald & Daniel M. Wolpert

2. Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK

James B. Heald, Máté Lengyel & Daniel M. Wolpert

3. Department of Neuroscience, Columbia University, New York, NY,  
USA

James B. Heald & Daniel M. Wolpert

4. Center for Cognitive Computation, Department of Cognitive Science,  
Central European University, Budapest, Hungary

Máté Lengyel

## Contributions

J.B.H. developed the model, implemented the model, performed the experiments, analysed the data and performed simulations. J.B.H. and D.M.W. designed the behavioural experiments. All of the authors were involved in the conceptualization of the study, developed techniques for analysing the model, interpreted results and wrote the paper.

## Corresponding author

Correspondence to [James B. Heald](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Additional details of the COIN model (related to Fig. 1). a-b, Hierarchy and generalization in contextual inference.

**a**, Local transition probabilities are generated in two steps via a hierarchical Dirichlet process. In the first step (top), an infinite set of global transition probabilities  $\beta$  are generated via a stochastic stick-breaking process ([Supplementary Information](#)). Probabilities are represented by the width of bar segments with different colours indicating different contexts. In the second step (bottom), for each context ('from context'), local transition probabilities to each context ('to context') are generated (a row of  $\{\Pi_i\}$ ) via a stochastic Dirichlet process and are equal to the global probabilities in expectation (bar a self-transition bias, which we set to zero here for clarity). (An analogous hierarchical Dirichlet process, not shown, is used to generate the global and local cue probabilities.) **b**, Contextual inference updates both the global and local transition probabilities. Context transition counts are maintained for all from-to pairs of known contexts and get updated based on the contexts inferred on two consecutive time points (responsibilities at time points  $t$  and  $t+1$ ). These updated context transition counts are used to update the inferred global transition probabilities  $\hat{\boldsymbol{\beta}}$ . The updated global transition probabilities and context transition counts produce new inferences about the inferred local transition probabilities  $\hat{\Pi}$ . Note that although the model infers full (Dirichlet) posterior distributions over both the global and local transition probabilities, for clarity here we only show the means of these posterior distributions (indicated by the hat notation). In the example shown, only row 3 of the context transition counts is updated (as context 3 has an overwhelming responsibility at time  $t$ ), but all rows of the local transition probabilities are updated due to the updating of the global transition probabilities (if the model were non-hierarchical, there would be no global transition probabilities, and so the local transition probabilities

would only be updated for context 3 via the updated context transition contexts). Thus, inferences about transition probabilities generalise from one context (here context 3) to all other contexts (here contexts 1 and 2) due to the hierarchical nature of the generative model. Note that when a novel context is encountered for the first time, its local transition probabilities are initialised based on  $\hat{\beta}$ , thus allowing well-informed inferences about transitions to be drawn immediately. **c–e**,

**Parameter inference in the COIN model for the simulation shown in Fig. 1c–h.** In addition to inferring states and contexts, the COIN model also infers context transition (**c**) and cue (**d**) probabilities, as well as the parameters of context-specific state dynamics (**e**). **c**, Transition probabilities. Top: Inferred global transition probabilities (solid lines) for transitioning into each known context (line colours) and the novel context (grey). Pale lines show inferred stationary probabilities for the same contexts, representing the expected proportion of time spent in each context given the current estimate of the local transition probabilities (below). Bottom three panels: inferred local transition probabilities from each context (colours as in top panel). Note that the local transition probability from context 1 to context 2 increases when cup 3 is handled (that is, when transitions from context 2 to itself are inferred to happen) due to the generalization of inferred transition probabilities across contexts. Also note that the local transition probabilities from context 3 are initialised based on the global transition probabilities (plus a self-transition bias).

**d**, Inferred global (top panel) and local cue probabilities for the three known contexts (bottom three panels) and cues (line colours). Although the model infers full (Dirichlet) posterior distributions over both transition (**c**) and cue probabilities (**d**), for clarity here we only show the means of these posterior distributions. **e**, Posterior distributions of the state drift (left) and retention parameters (right) for the three known contexts (colours as in **c**, novel context not shown for clarity). Although the model infers the joint distribution of the drift and retention parameters for each context, for clarity here we show the marginal distribution of each parameter separately. Note that drift and retention are inferred to be larger for the red context that is associated with the largest perturbation.

**Extended Data Fig. 2 Validation of the COIN model. a, Validation of the inference algorithm of the COIN model with a**

## single context.

We computed inferences in the COIN model with a single context based on synthetic observations (state feedback) generated by its generative model (Fig. 1a). Plots show the cumulative distributions of the posterior predictive p-values for the state variable (left) and the parameters governing its dynamics (retention, middle; drift, right). The posterior predictive p-value is computed by evaluating the cumulative distribution function of the model's posterior over the corresponding quantity at the true value of that quantity (as defined by the generative model). Empirical distributions of the posterior predictive p-values were collected across 4000 simulations (with different true retention and drift parameters), with 500 time steps in each simulation (during which the true state changes, but the true retention and drift parameters are constant). Note that although the true retention and drift parameters do not change during a simulation, inferences in the model about them still evolve in general, and so a new posterior predictive p-value is generated at each time step even for these quantities. If the model implements well-calibrated probabilistic inference under the correct generative model, the empirical distributions of the posterior predictive p-values should all be uniform. This is confirmed by all cumulative distributions of the posterior predictive p-values (orange and purple curves) approximating the identity line (thin black diagonal line). Orange curves show posterior predictive p-values under the corresponding marginals of the model's posterior. To give additional information about the model's joint posterior over the retention and drift parameters, we also show the cumulative distribution of the posterior predictive p-value for each parameter conditioned on the true value of the other parameter (retention | drift, and drift | retention, purple curves). **b, Validation of the inference algorithm of the COIN model with multiple contexts.** Simulations as in a but with additional synthetic observations (sensory cues) and multiple contexts allowed both during data generation and inference. Empirical distributions of the posterior predictive p-values were collected across 2000 simulations (with different true retention and drift parameters), with 500 time steps in each simulation (during which not only the true states change but also contexts transition, and sometimes novel contexts become active). Left column shows the true distributions of sensory cues, contexts and parameters. Inset shows the growth of the number of contexts over time

both during generation (blue) and inference (orange). Middle and right columns show the cumulative distributions of the posterior predictive p-values (pooled across data sets and time steps) for the observations (top row), contexts and state (middle row) and parameters (bottom row). To calculate the posterior predictive p-values for the context, inferred contexts were relabelled by minimising the Hamming distance between the relabelled context sequence and the true context sequence ([Supplementary Information](#)). For the parameters, the posterior predictive p-values were calculated with respect to both the marginal distributions (retention and drift) and the conditional distributions (retention | drift, and drift | retention) as in **a**. The cumulative probability curves approximate the identity line (thin black diagonal line) showing that the inferred posterior probability distributions are well calibrated. **c, Parameter recovery in the COIN model related to Fig. 2**. Plots show the COIN model parameters that were recovered (y-axes) from fits to 10 synthetic data sets generated with the COIN model parameters (true, x-axes) obtained from the fits to each participant in the spontaneous ( $n = 8$ ) and evoked ( $n = 8$ ) recovery experiments (Extended Data Fig. [3](#)). Vertical bars show the interquartile range of the recovered parameters for each participant. While several parameters are recovered with good accuracy  $\{(\{\sigma\}_{\rm rm}\{q\}), \{\mu\}_{\rm rm}\{a\}, \{\sigma\}_{\rm rm}\{d\}, \{\sigma\}_{\rm rm}\{m\})\}$ , others are not ( $\alpha$ , and in particular  $\{\sigma\}_{\rm rm}\{a\}$ ) and  $(\rho)$ . We expect that with richer paradigms and larger data sets, all parameters would be recovered accurately. Most importantly, despite partial success with recovering individual parameters, model recovery shows that recovered parameter sets taken as a whole can still be used to accurately identify whether data was generated by the dual-rate or COIN model (**d**). Note that we make no claims about individual parameters in this study as our focus is on model class recovery. **d-e, Model recovery for spontaneous (d) and evoked recovery experiments (e) related to Fig. 2**. Synthetic data sets were generated using one of two models (COIN model, cyan; dual-rate model, green). Parameters used for each model were those obtained from the fits to each participant in the spontaneous ( $n = 8$ ) and evoked ( $n = 8$ ) recovery experiments (Extended Data Fig. [3](#)), that is, for the COIN model, these were the same synthetic data sets as those used in **c**. Then the same model comparison method that we used on real data (Fig. [2c](#), **e**, insets) was used to recover the model that generated each synthetic data

set ([Methods](#)). Arrows connect true models (used to generate synthetic data, disks on top) to models that were recovered from their synthetic data (pie-chart disks at bottom). Arrow colour indicates identity of recovered model, arrow thickness and percentages indicate probability of recovered model given true model. Bottom disk sizes and pie-chart proportions show total probability of recovered model and posterior probability of true model given recovered model (assuming a uniform prior over true models), respectively, with percentages specifically indicating posterior probability of the correct model. These results show that the model recovery process is generally very accurate and actually biased against the COIN model in favour of the dual-rate model.

### [Extended Data Fig. 3 COIN model parameters.](#)

Left column: Parameters for illustrating the COIN model (**I**: purple), model validation (**V**: brown) and fits to individuals in the spontaneous (**S**: blue) and evoked (**E**: green) recovery experiments, to the average of both groups (**A**: cyan), and individuals in the memory-updating experiment (**M**: red).

Right: scatter plots for all pairs of parameters for the six groups. The overlap of data points suggest parameters are similar across experiments.  $\sigma_{\text{rm}\{q\}}$ : process noise s.d. (equation 3);  $\mu_{\text{rm}\{a\}}$ ,  $\sigma_{\text{rm}\{a\}}$ : prior mean and s.d. for context-specific state retention factors (equation 10);  $\sigma_{\text{rm}\{d\}}$ : prior s.d. for context-specific state drifts (equation 10);  $\alpha$ : concentration of local transition probabilities (equation 8);  $\rho$ : self-transition bias parameter (equation 18);  $\sigma_{\text{rm}\{m\}}$ : motor noise s.d. (equation 17);  $\alpha^{\text{rm}\{e\}}$ : concentration of local cue probabilities (equation 9). Parameters used in the figures is as follows.

**I**: Fig. 1 and Extended Data Fig. 1c–e. **V**: Extended Data Fig. 2a, b. **S**: Fig. 2c, Extended Data Fig. 6f (column 1) and Extended Data Fig. 2d. **E**: Fig. 2e, Extended Data Fig. 6f (column 3) and Extended Data Fig. 2e. **S & E**: Extended Data Fig. 2c. **A**: Fig. 2b and d, Extended Data Fig. 5 and Extended Data Fig. 9 (bias added for visuomotor rotation experiments: Extended Data Fig. 5a–j, p–s and Extended Data Fig. 9e–l). **M**: Fig. 3 and Extended Data Fig. 7a–d. **S, E & M**: (all parameters, but  $\alpha^{\text{rm}\{e\}}$ ): Fig. 4 and Extended Data Fig. 8. The robustness analyses (Extended Data Fig. 4) used perturbed versions of the same parameters as

the corresponding unperturbed simulations. To reduce the number of free parameters in the model, we set the parameters of the hierarchical Dirichlet process that determine the expected effective number of contexts or cues,  $\gamma$  (equation 7) and  $(\{\gamma\}^{\{\rm{e}\}})$  (equation 9), respectively, both to 0.1, the prior mean for context-specific state drifts  $(\{\mu\}_{\{\rm{d}\}})$  to zero (equation 10) and the standard deviation of the sensory noise  $(\{\sigma\}_{\{\rm{s}\}})$  to 0.03 when fitting or simulating the model, with the variance of the observation noise (equations 5 and 19) set to  $(\{\sigma\}_{\{\rm{r}\}})^2 = (\{\sigma\}_{\{\rm{s}\}})^2 + (\{\sigma\}_{\{\rm{m}\}})^2$ . For visuomotor rotation experiments (Extended Data Fig. 5a–j, p–s and Extended Data Fig. 9e–l), we set the mean of the prior of the bias  $(\{\mu\}_{\{\rm{b}\}})$  to zero (equation 20) and its s.d.  $(\{\sigma\}_{\{\rm{b}\}})$  to  $70^{-1}$ .

## Extended Data Fig. 4 Robustness analysis of the main COIN model results.

To test how robust the behaviour of the COIN model is, we added noise to the parameters fit to the individual participants in the spontaneous recovery, evoked recovery and memory updating experiments and re-simulated the paradigms in Figs. 2–4: spontaneous recovery (**a**), evoked recovery (**b**), memory updating (**c**), savings (**d**), anterograde interference (**e**) and environmental consistency (**f**). For each experiment, we simulated the COIN model for the same participants as in Figs. 2–4 but perturbed each participant's parameter values. That is, for each parameter (suitably transformed to be unbounded) we calculated the standard deviation across participants (relevant for the given paradigm or set of paradigms) and then perturbed each participant's (transformed) parameter by zero-mean Gaussian noise whose standard deviation was a fraction ( $\lambda = 0, 0.05, 0.5$  or  $1.0$ ) of this empirical standard deviation, after which we used the inverse transform to obtain the actual parameter used in these perturbed simulations. For parameters that are constrained to be non-negative ( $(\{\sigma\}_{\{\rm{q}\}})$ ,  $(\{\sigma\}_{\{\rm{a}\}})$ ,  $(\{\sigma\}_{\{\rm{d}\}})$ ,  $(\{\alpha\})$ ,  $(\{\alpha\}^{\{\rm{e}\}})$ ,  $(\{\sigma\}_{\{\rm{m}\}})$ ), we used a logarithmic transformation, whereas for parameters constrained to be on the unit interval ( $(\{\mu\}_{\{\rm{a}\}})$ ,  $(\rho)$ ), we used a logit transformation. Column 1: experimental data

(plotted as in Figs. 2–4). Columns 2–5: output of the COIN model for different amounts of noise added to the parameters. Note that the simulations were not conditioned on the actual adaptation data of individual participants (in contrast to the original simulations of Figs. 2 and 3) because these data are not available for the experiments shown in Fig. 4 (for which the original simulations were already performed using this ‘open-loop’ simulation approach). The robustness analysis shows that most predictions of the COIN model are robust to changes in the parameters and only start to deviate for large parameter changes ( $\lambda = 1$ ) in some of their quantitative details (such as the magnitude of spontaneous recovery). Note that  $\lambda = 1$  leads to changes in parameters that are of the same magnitude as randomly shuffling the parameters across participants.

**Extended Data Fig. 5 History dependence of contextual inference. a-j, Contextual inference underlies the elevated level of spontaneous recovery after ‘overlearning’.**

**a**, Spontaneous recovery paradigm for visuomotor learning in which the length of the exposure ( $P^+$ ) phase is tripled from 200 trials (‘standard’ paradigm, pink) to 600 trials (‘overlearning’ paradigm, green). For comparison, paradigms are aligned to the end of the exposure phase. **b**, Adaptation in the COIN model for the standard and overlearning paradigms (same parameters as in Fig. 2b and d but with the addition of a bias parameter; see [Supplementary Information](#) and also Extended Data Fig. 3, parameter set A). Adaptation corresponds to reach angle normalized by the size of the experimentally imposed visuomotor rotation. Note elevated level of spontaneous recovery after overlearning compared to the standard paradigm, qualitatively matching visuomotor learning data in Fig. 4a of Ref. 13. **c-f**, Internal representations of the COIN model for the standard paradigm. Inferred bias (**c**) and predicted state (**d**) distributions for each context (colours). **e**, Predicted probabilities of each context (with zoomed view starting from near the end of  $\backslash(\{\{\rm P\}\}\}^{\wedge\{+\}}\backslash)$  exposure), colours as in c-d, grey is novel context as in Fig. 1f. **f**, Predicted state feedback (predicted state plus bias) distribution (purple), which is a mixture of the individual contexts’ predicted state feedback distributions (not shown) weighted by their predicted probabilities (**e**). Total adaptation (cyan line) is

the mean of the predicted state feedback distribution. **g-j**, same as **c-f** for the overlearning paradigm. For comparison, the dashed horizontal lines in both paradigms show the final level of each variable for the red context in the standard paradigm. Note that overlearning leaves inferences about biases and states largely unchanged (compare 1 in **c & g** and 2 in **d & h**) but leads to higher predicted probabilities of the  $\langle \{ \rm{P} \} \rangle^{\{+\}}$  context (red) in the channel-trial phase (compare 3 in **e & i**), reflecting the true statistics of the experiment in which  $\langle \{ \rm{P} \} \rangle^{\{+\}}$  occurred more frequently. In turn, this makes the  $\langle \{ \rm{P} \} \rangle^{\{+\}}$  bias and state contribute more to total adaptation in the channel-trial phase, thus explaining higher levels of spontaneous recovery. Therefore, differences between conditions are explained by contextual inference rather than by differences in bias or state inferences. The results are qualitatively similar when simulated as a force-field paradigm (that is, without bias, not shown). **k-o**, **Contextual inference underlies reduced spontaneous recovery following pre-training with**  $\langle \{ \rm{P} \} \rangle^{\{-}}$ . **k**, Adaptation in the channel-trial phase of a typical spontaneous recovery paradigm (standard, pink, as in Fig. 2b) and two modified versions of the paradigm in which the  $\langle \{ \rm{P} \} \rangle^{\{+\}}$  phase is preceded by a  $\langle \{ \rm{P} \} \rangle^{\{-}}$  (pre-training) phase in which  $\langle \{ \rm{P} \} \rangle^{\{-}}$  is either introduced and removed abruptly ( $\langle \{ \rm{P} \} \rangle_{-} \langle \{ \rm{abrupt} \} \rangle^{\{-}}$ , dark green) or gradually ( $\langle \{ \rm{P} \} \rangle_{-} \langle \{ \rm{g} \} \rangle \langle \rm{r} \rangle \langle \rm{a} \rangle \langle \rm{d} \rangle \langle \rm{u} \rangle \langle \rm{a} \rangle \langle \rm{l} \rangle^{\{-}}$ , light green). Data reproduced from Ref. 14. **l-o**, Simulation of the COIN model for the same paradigms (same parameters as in Fig. 2b and d; Extended Data Fig. 3, parameter set A), plotted as in Fig. 2b-c. In each paradigm, contexts are coloured according to their order of instantiation during inference (blue  $\rightarrow$  red  $\rightarrow$  orange). Note that pre-training with  $\langle \{ \rm{P} \} \rangle^{\{-}}$  (either abrupt or gradual) leaves inferences about states within each context largely unchanged at the beginning of the channel-trial phase (compare corresponding numbers 1-2 in column 2 across **m-o**). However, the pre-training leads to higher predicted probabilities of the  $\langle \{ \rm{P} \} \rangle^{\{-}}$  context initially (compare number 3 in **m** to number 3 in **n & o**) and throughout the channel-trial phase (compare number 4 across **m-o**) reflecting the true statistics of the experiment in which  $\langle \{ \rm{P} \} \rangle^{\{-}}$  occurred more frequently (compare column 1 across **m-o**). In turn, this makes the  $\langle \{ \rm{P} \} \rangle^{\{-}}$  state contribute more to total adaptation, thus explaining the reduction in both the

initial and final levels of adaptation during the channel-trial phase in the  $\{\{\rm{P}\}\}^{\pm 1}$  and  $\{\{\rm{P}\}\}^{\pm 2}$  groups. Therefore, as in Fig. 4, differences between conditions are explained by contextual inference rather than state inference. **p-s, Contextual inference underlies slower de-adaptation following a gradually introduced perturbation.** p, Adaptation (normalized reach angle, as in b) in a paradigm in which a visuomotor rotation is introduced abruptly (pink) or gradually (green) and then removed abruptly. Data reproduced from Ref. 17. q-s, Simulation of the COIN model on the abrupt (q, pink, and r) and gradual (q, green, and s) paradigms (same parameters as in Fig. 2b and d but with the addition of a bias parameter; Extended Data Fig. 3, parameter set A) plotted as in b-j. Note that contexts are coloured according to their order of appearance during inference (blue \to red). In response to the abrupt introduction of the  $\{\{\rm{P}\}\}^{\pm 1}$  perturbation, a new memory is created (1). In contrast, the gradual introduction of the  $\{\{\rm{P}\}\}^{\pm 1}$  perturbation prevents the creation of a new memory, thus requiring changes in the inferred bias and state of the original memory associated with  $\{\{\rm{P}\}\}^0$  (2, blue context) to account for the slowly increasing perturbation. Therefore, the ‘blue’ context is inferred to be active throughout the exposure phase (3) and becomes associated with a  $\{\{\rm{P}\}\}^{\pm 1}$ -like state. However, at the beginning of the abruptly introduced post-exposure ( $P^0$ ) phase, a new memory is created (4), which has a low initial predicted probability that can only be increased by repeated experience with  $P^0$  (5). This leads to slower de-adaptation in the post-exposure phase compared to the abrupt paradigm (6), in which the original context associated with  $\{\{\rm{P}\}\}^0$  (blue) is protected (7) and can be reinstated quickly (8) as the  $\{\{\rm{P}\}\}^0$  local self-transition probability has been learned to be higher during the pre-exposure phase. Note that the smaller errors caused by the gradual perturbation relative to the abrupt condition are better accounted for by an error in the state rather than an error in the bias, and therefore the state is updated more than the bias. The results are qualitatively similar when simulated as a force field paradigm (that is, without bias, not shown).

## Extended Data Fig. 6 Additional analyses of spontaneous and evoked recovery related to Fig. 2. a-c, Mathematical analysis of spontaneous and evoked recovery.

The channel-trial phase of spontaneous recovery and evoked recovery (after the two  $\{\{ \rm P \} \}^{+}$  trials) simulated in a simplified setting ([Supplementary Information](#)) with two contexts that are initialized to have equal but opposite state estimates (**a**) and equal (spontaneous recovery, solid) or highly unequal (evoked recovery, dashed) predicted probabilities (**b**). For the two contexts, the retention parameters are assumed to be constant and equal, and the drift parameters are assumed to be constant, of the same magnitude but opposite sign. Mean adaptation (**c**), which in the COIN model is the average of the state estimates (**a**) weighted by the corresponding predicted probabilities (**b**), shows the classic pattern of spontaneous recovery (solid, cf. Fig. [2b, c](#)) and the characteristic abrupt rise of evoked recovery (dashed, cf. Fig. [2d, e](#)). Note that although in the full model, state estimates are different between evoked and spontaneous recovery following the two  $\{\{ \rm P \} \}^{+}$  trials, here we assume they are the same (no separate solid and dashed lines in **a**) for simplicity and to demonstrate that the difference in mean adaptation between the two paradigms (**c**) can be accounted for by differences in contextual inference alone (**b**, cf. Fig. [2b and d](#), top right). Circles on the right show steady-state values of inferences and adaptation. Note that in both paradigms, adaptation is predicted to decay to a non-zero asymptote (see also **e**). **d, State-space model fits to adaptation data from the spontaneous and evoked recovery groups.** Solid lines show the mean fits across participants of the two-state model (5 parameters, top row) and the three-state model (7 parameters, bottom row) to the spontaneous recovery (left column) and evoked recovery (right column) data sets. Mean  $(\pm \text{s.e.m.})$  adaptation on channel trials shown in black (same as in Fig. [2c and e](#)). Insets show differences in BIC (nats) between the two-state model and the three-state model for individual participants (positive values in green indicate evidence in favour of the two-state model, and negative values in purple indicate evidence in favour of the three-state model). At the group level, the two-state model was far superior to the three-state model ( $(\Delta)$  group-level BIC of 64.2 and 78.4 nats in favour of the two-state model for the spontaneous and evoked recovery groups, respectively). Individual states

are shown for the two-state model (top, blue and red). Both the fast and slow processes adapt to  $(\{\rm{P}\})^+$  during the extended initial learning period. The  $(\{\rm{P}\})^-$  phase reverses the state of the fast process, but not of the slow process, so that they cancel when summed resulting in baseline performance. Spontaneous recovery during the  $(\{\rm{P}\})^c$  phase is then explained by the fast process rapidly decaying, revealing the state of the slow process that has remained partially adapted to  $(\{\rm{P}\})^+$ . Note that this explanation arises because in multi-rate models all processes contribute equally to the motor output at all times. This is fundamentally different from the expression and updating of multiple context-specific memories in the COIN model, which are dynamically modulated over time according to ongoing contextual inference.

**e, Evoked recovery does not decay exponentially to zero.** According to the COIN model, adaptation in the channel-trial phase of evoked recovery can be approximated by exponential decay to a non-zero (positive) asymptote (**a-c**, Fig. 2e, [Supplementary Information](#)). To test this prediction, we fit an exponential function that either decays to zero (light and dark green) or decays to a non-zero (constrained to be positive) asymptote (cyan) to the adaptation data of individual participants in the evoked recovery group after the two  $(\{\rm{P}\})^+$  trials (black arrow). The two zero-asymptote models differ in terms of whether they are constrained to pass through the datum on the first trial (light green) or not (dark green). The mean fits across participants for the models that decay to zero (green) fail to track the mean adaptation (black,  $\pm$  s.e.m. across participants), which shows an initial period of decay followed by a period of little or no decay. The mean fit for the model that decays to a non-zero asymptote (cyan) tracks the mean adaptation well and was strongly favoured in model comparison ( $\Delta$  group-level BIC of 944.3 and 437.7 nats compared to the zero-asymptote fits with constrained and unconstrained initial values, respectively). Note that fitting to individual participants excludes the confound of finding a more complex time course (e.g. one with a non-zero asymptote) only due to averaging across participants that each show a different simple time course (e.g. all with zero asymptote but different time constants).

**f, COIN and dual-rate model fits for individual participants in the spontaneous and evoked recovery groups.** Data and model predictions are shown for individual participants as in Fig. 2c and e for across-participant averages.

Participants in the **S** and **E** groups are ordered by decreasing BIC difference between the dual-rate and COIN model (that is, **S1**'s and **E1**'s data most favour the COIN model), as in insets of Fig. [2c and e](#). Note that the COIN model can account for much of the heterogeneity of spontaneous recovery (e.g. from large in **S1** to minimal in **S6**) and evoked recovery (e.g. from large in **E1** to minimal in **E7**).

**Extended Data Fig. 7 Additional analyses of the memory updating experiment (related to Fig. 3). a-b, Memory updating experiment: time-course of learning.**

**a**, Adaptation on channel trials at the end of each block of force-field trials in the training phase (purple), which occur before  $\{\{\rm{P}\}\}^0$  washout trials, and on the first channel trial of triplets within each block (orange), which occurs after  $\{\{\rm{P}\}\}^0$  washout trials. Data is mean  $(\pm)$  s.e.m. across participants and lines show mean of COIN model fits (8 parameters, Extended Data Fig. [3](#)). **b**, Single-trial learning on triplets that were consistent with the training contingencies. Data (mean  $(\pm)$  s.e.m. across participants) with mean of COIN model fits across participants. Positive learning reflects changes in the direction expected based on the force field of the exposure trial (an increase following  $\{\{\rm{P}\}\}^+$  and a decrease following  $\{\{\rm{P}\}\}^-$ ). **c-d,** **Mathematical analysis of single-trial learning.** Single-trial learning in the COIN model (column 1) for the four cue–perturbation triplets in the pre-training phase (**c**) and the post-training phase (**d**) in the memory updating experiment. The COIN model was fit to each participant and model fits are shown as mean  $(\pm)$  s.e.m. (single-trial learning, full model prediction) or mean (dot product, posterior, prior and likelihood) across  $n = 24$  participants. Single-trial learning (column 1) is approximately proportional to a dot product (column 2) between the vector of posterior context probabilities (responsibilities) on the exposure trial of the triplet and the vector of predicted context probabilities on the subsequent channel trial (see the [Supplementary Information](#) for derivation). This dot product can be further approximated by collapsing the vector of predicted probabilities to a one-hot vector, that is, by the responsibility  $(p(c_t=c^*)|q_t, y_t, \dots)$  (column 3) of the context that is predominantly expressed on the second channel trial of the triplet ( $c^*$ , the context with the

highest predicted probability on the second channel trial of the triplet), where  $\dots$  denotes all observations before time  $t$  (as in Fig. 1). This responsibility is proportional to a product of two terms. The first term is the prior context probability  $(p(c_t = c | \text{last}, q_t, \dots))$  (column 4), that is, the predicted context probability before experiencing the perturbation (as in Fig. 1f), which is already conditioned on the sensory cue visible from the outset of the trial. The second term expresses the likelihood of the state feedback in that context  $(p(y_t | c_t = c, \dots))$  (column 5). Because prior to learning neither cues nor feedback are yet consistently associated with a particular context, the COIN model predicts that the prior and likelihood, and thus total single-trial learning should all be largely uniform across contexts before training. **e-f, The effects of cue and perturbation on single-trial learning in individual participants.** **e,** Single-trial learning (post-training) shown as a function of perturbation separated by cue (left) or as a function of cue separated by perturbation (right) for each participant (lines). Note a significant effect for both the perturbation and the cue. **f,** Scatter plot of cue effect ( $(\{\rm P\}_1^{(+)} - \{\rm P\}_1^{(-)})$ ) against perturbation effect ( $(\{\rm P\}_2^{(+)} - \{\rm P\}_2^{(-)})$ ) for each participant (dots). Solid lines show medians of corresponding effects. Note the lack of anti-correlation between two effects.

### Extended Data Fig. 8 Additional analysis of the effect of environmental consistency on single-trial learning related to Fig. 4c.

Columns 1 & 2: experimental paradigm and data replotted from Ref. 5. Participants experienced repeating cycles of  $(\{\rm P\}^+)$  trials of varying lengths (column 1: 20  $(\{\rm P\}^+)$  trials in P20, 7 in P7, 1 in P1 and 1 followed by 1  $(\{\rm P\}^-)$  trial in P1N1) in between  $(\{\rm P\}^0)$  trials. To assess single-trial learning (column 2) during exposure to the environments, channel trials were randomly interspersed before and after the first  $(\{\rm P\}^+)$  trial in a subset of the force-field cycles. Columns 3 to 5 show the output and internal inferences of the COIN model in the same format as Fig. 4c (same parameters as in Fig. 4;

Extended Data Fig. 3, parameter set S, E & M). The COIN model qualitatively reproduced the pattern of changes in single-trial learning seen over repeated cycles in this paradigm. As in Fig. 4, differences in the apparent learning rate were not driven by differences in either the proper learning rate (Kalman gain) or the underlying state (column 4) but were instead driven by changes in contextual inference (column 5).

### Extended Data Fig. 9 Cognitive processes and the COIN model.

#### a-d, Maintenance of context probabilities may require working memory.

**a**, Adaptation in a spontaneous recovery paradigm in which a non-memory (pink) or working memory task (green) is performed at the end of the \(\{\{\rm P\}\}^{\{-\}}\}\) phase before starting the channel-trial phase (data reproduced from Ref. 22). Initial adaptation in the channel-trial phase (inset) shows the working memory task abolishes spontaneous recovery and leads to adaptation akin to evoked recovery (cf. Extended Data Fig. 6a–c). **b-d**, COIN model simulation in which the working memory task abolishes the (working) memory of the context responsibilities on the last trial of the \(\{\{\rm P\}\}^{\{-\}}\}\) phase but not the context transition (and thus stationary) probabilities (same parameters as in Fig. 2b and d; Extended Data Fig. 3, parameter set A), plotted as in Fig. 2b, c. The circles on the predicted probability (zoomed view) show the values on the first trial in the channel-trial phase. **d**, as (c) but for the working memory task. The predicted probabilities on the first trial in the channel-trial phase are set to the values under the stationary distribution (shown on every trial in the simulation of Extended Data Fig. 1c). We calculate the stationary context distribution by solving \(\{\boldsymbol{\psi}\} = \{\boldsymbol{\psi}\} \hat{\boldsymbol{P}}\}\) for \(\{\boldsymbol{\psi}\}\) (a row vector), subject to the constraint that \(\{\boldsymbol{\psi}\}\) is a valid probability distribution (i.e. all elements of \(\{\boldsymbol{\psi}\}\) are non-negative and sum to 1), where \(\{\hat{\boldsymbol{P}}\}\) is the expected local transition probability matrix. **e-l**, **Explicit versus implicit learning in the COIN model.** **e**, Results of a spontaneous recovery paradigm (as in Fig. 2b) for visuomotor learning. Adaptation is computed as participants' reach angle normalized by the size of the experimentally imposed visuomotor rotation. Explicit learning (dark green) is measured by participants indicating their intended reach direction.

Implicit learning (light green) is obtained as the difference between total adaptation (solid pink) and explicit learning. In the visual error-clamp phase ( $\langle \{ \{ \backslash \mathrm{rm\{P\}} \} \}^{\wedge} \{ \{ \backslash \mathrm{mathsf\{c\}} \} \} \rangle$ ), participants were told to stop using any aiming strategy so that the direction they moved was taken as the implicit component of learning. A control experiment (dashed pink) was also performed in which there was no reporting of intended reach direction. Data reproduced from Ref. [24](#). **f-l**, Simulation of the COIN model on the same paradigm (same parameters as in Fig. [2b](#) and d but with the addition of a bias parameter; Extended Data Fig. [3](#), parameter set **A**). **b**, Predictions for experimentally observable quantities. Light green line: implicit learning is the average bias across contexts weighted by the predicted probabilities (cyan line in **j**). Dark green line: explicit learning is the state of the most responsible context on the previous trial (black line in **h**). Solid pink line: total adaptation for the reporting condition is the sum of explicit and implicit learning (as in **e**). Dashed pink line: total adaptation for the non-reporting condition is the average predicted state feedback across contexts weighted by the predicted probabilities (cyan line in **l**, as in all experiments that had no reporting element). **g-h**, Inferred bias (**g**) and predicted state (**h**) distributions for each context (colours), with black line showing the mean state of the most responsible context (coloured line below axis) for trials on which an explicit report was solicited. **i**, Predicted probability of each context. Colours as in **g-h**, grey is novel context as in Fig. [1f](#). **j-k**, Inferred bias (**j**) and predicted state (**k**) distributions (purple), obtained as mixtures of the respective distributions of individual contexts (**g-h**) weighted by their predicted probabilities (**i**), and their means (cyan lines). **l**, Predicted state feedback distribution (purple, computed as the sum of bias in **j** and predicted state in **k**) and its mean (cyan).

## Extended Data Table 1 Comparison of the COIN model to other models

## Supplementary information

### Supplementary Information

Supplementary Information 1–8. See the Supplementary Information contents page for details.

## Reporting Summary

## Peer Review File

# Rights and permissions

## Reprints and Permissions

# About this article

## Cite this article

Heald, J.B., Lengyel, M. & Wolpert, D.M. Contextual inference underlies the learning of sensorimotor repertoires. *Nature* **600**, 489–493 (2021).  
<https://doi.org/10.1038/s41586-021-04129-3>

- Received: 08 December 2020
- Accepted: 13 October 2021
- Published: 24 November 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04129-3>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

# Further reading

- **Context is key for learning motor skills**

- Anne G. E. Collins
- Samuel D. McDougle

*Nature* (2021)

## **Context is key for learning motor skills**

- Anne G. E. Collins
- Samuel D. McDougle

News & Views 24 Nov 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04129-3>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 08 December 2021](#)

# Exercise plasma boosts memory and dampens brain inflammation via clusterin

- [Zurine De Miguel](#)<sup>1,2,3 nAff11</sup>,
- [Nathalie Khoury](#)<sup>1,2,3 na1</sup>,
- [Michael J. Betley](#)<sup>1,4 na1</sup>,
- [Benoit Lehallier](#) [ORCID: orcid.org/0000-0001-7452-3785](#)<sup>1,2,3 nAff12</sup>,
- [Drew Willoughby](#)<sup>1,2,3</sup>,
- [Niclas Olsson](#) [ORCID: orcid.org/0000-0002-9746-0542](#)<sup>5 nAff13</sup>,
- [Andrew C. Yang](#) [ORCID: orcid.org/0000-0002-6756-4746](#)<sup>1,2,3</sup>,
- [Oliver Hahn](#)<sup>1,2,3</sup>,
- [Nannan Lu](#)<sup>1,2,3</sup>,
- [Ryan T. Vest](#)<sup>1,2,3</sup>,
- [Liana N. Bonanno](#)<sup>1,2,3</sup>,
- [Lakshmi Yerra](#)<sup>6</sup>,
- [Lichao Zhang](#) [ORCID: orcid.org/0000-0003-1811-7568](#)<sup>7</sup>,
- [Nay Lui Saw](#)<sup>8</sup>,
- [J. Kaci Fairchild](#)<sup>6</sup>,
- [Davis Lee](#)<sup>1,2,3</sup>,
- [Hui Zhang](#) [ORCID: orcid.org/0000-0001-6787-7403](#)<sup>1,2,3</sup>,
- [Patrick L. McAlpine](#) [ORCID: orcid.org/0000-0002-9950-7156](#)<sup>9</sup>,
- [Kévin Contrepois](#)<sup>10</sup>,
- [Mehrdad Shamloo](#)<sup>8</sup>,
- [Joshua E. Elias](#)<sup>5,7</sup>,
- [Thomas A. Rando](#) [ORCID: orcid.org/0000-0001-5843-8564](#)<sup>1,2,6</sup> &
- [Tony Wyss-Coray](#) [ORCID: orcid.org/0000-0001-5893-0831](#)<sup>1,2,3</sup>

[Nature](#) volume **600**, pages 494–499 (2021)

- 21k Accesses
- 1133 Altmetric
- [Metrics details](#)

## Subjects

- [Adult neurogenesis](#)
- [Molecular neuroscience](#)
- [Neuroimmunology](#)

## Abstract

Physical exercise is generally beneficial to all aspects of human and animal health, slowing cognitive ageing and neurodegeneration<sup>1</sup>. The cognitive benefits of physical exercise are tied to an increased plasticity and reduced inflammation within the hippocampus<sup>2,3,4</sup>, yet little is known about the factors and mechanisms that mediate these effects. Here we show that ‘runner plasma’, collected from voluntarily running mice and infused into sedentary mice, reduces baseline neuroinflammatory gene expression and experimentally induced brain inflammation. Plasma proteomic analysis revealed a concerted increase in complement cascade inhibitors including clusterin (CLU). Intravenously injected CLU binds to brain endothelial cells and reduces neuroinflammatory gene expression in a mouse model of acute brain inflammation and a mouse model of Alzheimer’s disease. Patients with cognitive impairment who participated in structured exercise for 6 months had higher plasma levels of CLU. These findings demonstrate the existence of anti-inflammatory exercise factors that are transferrable, target the cerebrovasculature and benefit the brain, and are present in humans who engage in exercise.

This is a preview of subscription content

## **Access options**

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

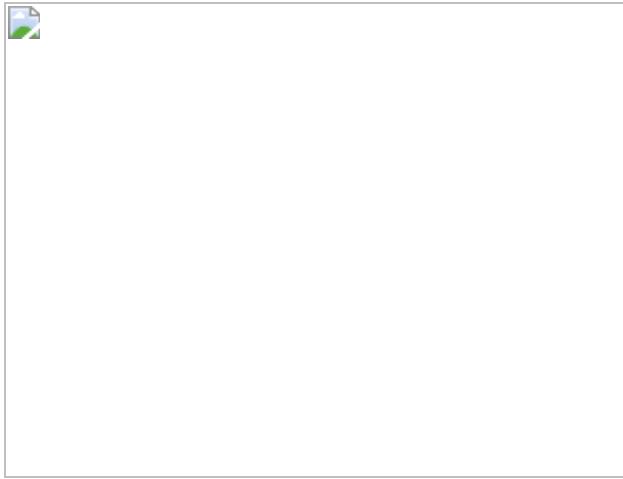
[Rent or Buy](#)

All prices are NET prices.

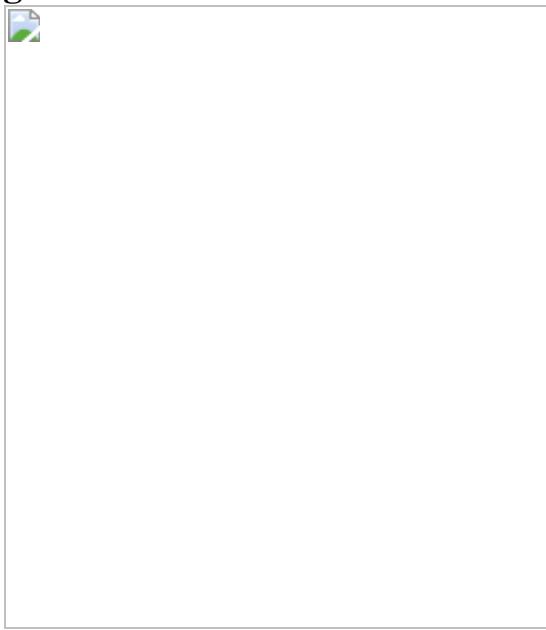
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

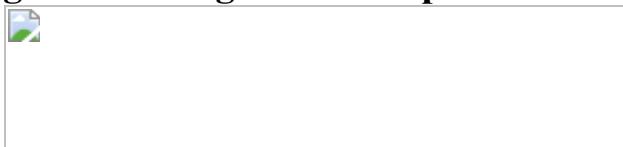
**Fig. 1: RP induces neuroplasticity, improves cognition and reduces inflammation.**



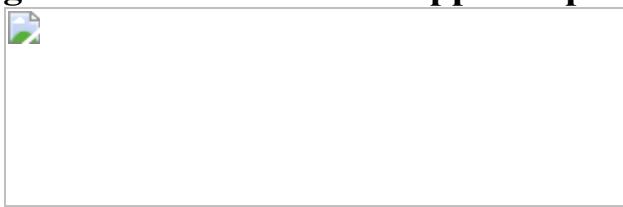
**Fig. 2: RP counteracts LPS-induced neuroinflammation.**



**Fig. 3: Running alters complement and coagulation proteins.**



**Fig. 4: Clusterin reduces hippocampal inflammation.**



## Data availability

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>55</sup> partner repository under dataset identifier [PXD022262](#) and [PXD027406](#) for male and female mouse plasma proteins, respectively. Single-cell data and bulk RNA-seq datasets have been deposited at the Gene Expression Omnibus under accession number [GSE164401](#). [Source data](#) are provided with this paper.

## Code availability

Data analyses and graphing have been carried out using free available software packages. When appropriate, custom code from previous literature was cited in the text and is available from the corresponding authors on request.

## References

1. 1.  
Prakash, R. S., Voss, M. W., Erickson, K. I. & Kramer, A. F. Physical activity and cognitive vitality. *Annu. Rev. Psychol.* **66**, 769–797 (2015).
2. 2.  
Gleeson, M. et al. The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease. *Nat. Rev. Immunol.* **11**, 607–615 (2011).
3. 3.  
He, X. F. et al. Voluntary exercise promotes glymphatic clearance of amyloid beta and reduces the activation of astrocytes and microglia in aged mice. *Front. Mol. Neurosci.* **10**, 144 (2017).
4. 4.

van Praag, H., Christie, B. R., Sejnowski, T. J. & Gage, F. H. Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proc. Natl Acad. Sci. USA* **96**, 13427–13431 (1999).

5. 5.

Hawley, J. A., Hargreaves, M., Joyner, M. J. & Zierath, J. R. Integrative biology of exercise. *Cell* **159**, 738–749 (2014).

6. 6.

Chin, L. M., Keyser, R. E., Dsurney, J. & Chan, L. Improved cognitive performance following aerobic exercise training in people with traumatic brain injury. *Arch. Phys. Med. Rehabil.* **96**, 754–759 (2015).

7. 7.

Horder, H. et al. Midlife cardiovascular fitness and dementia: a 44-year longitudinal population study in women. *Neurology* **90**, e1298–e1305 (2018).

8. 8.

da Costa Daniele, T. M. et al. Exercise effects on brain and behavior in healthy mice, Alzheimer’s disease and Parkinson’s disease model—a systematic review and meta-analysis. *Behav. Brain Res.* **383**, 112488 (2020).

9. 9.

Trejo, J. L., Carro, E. & Torres-Aleman, I. Circulating insulin-like growth factor I mediates exercise-induced increases in the number of new neurons in the adult hippocampus. *J. Neurosci.* **21**, 1628–1634 (2001).

10. 10.

Fabel, K. et al. VEGF is necessary for exercise-induced adult hippocampal neurogenesis. *Eur. J. Neurosci.* **18**, 2803–2812 (2003).

11. 11.

Leiter, O. et al. Exercise-induced activated platelets increase adult hippocampal precursor proliferation and promote neuronal differentiation. *Stem Cell Rep.* **12**, 667–679 (2019).

12. 12.

Horowitz, A. M. et al. Blood factors transfer beneficial effects of exercise on neurogenesis and cognition to the aged brain. *Science* **369**, 167–173 (2020).

13. 13.

Chen, M. B. et al. Brain endothelial cells are exquisite sensors of age-related circulatory cues. *Cell Rep.* **30**, 4418–4432 (2020).

14. 14.

Batista, C. R. A., Gomes, G. F., Candelario-Jalil, E., Fiebich, B. L. & de Oliveira, A. C. P. Lipopolysaccharide-induced neuroinflammation as a bridge to understand neurodegeneration. *Int. J. Mol. Sci.* **20**, 2293 (2019).

15. 15.

Conway, E. M. Complement-coagulation connections. *Blood Coagul. Fibrinolysis* **29**, 243–251 (2018).

16. 16.

Markiewski, M. M., Nilsson, B., Ekdahl, K. N., Mollnes, T. E. & Lambris, J. D. Complement and coagulation: strangers or partners in crime? *Trends Immunol.* **28**, 184–192 (2007).

17. 17.

Hicks, A. L., Kent-Braun, J. & Ditor, D. S. Sex differences in human skeletal muscle fatigue. *Exerc. Sport Sci. Rev.* **29**, 109–112 (2001).

18. 18.

Yip, K. S., Suvorov, A., Connerney, J., Lodato, N. J. & Waxman, D. J. Changes in mouse uterine transcriptome in estrus and proestrus. *Biol. Reprod.* **89**, 13 (2013).

19. 19.

Harold, D. et al. Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).

20. 20.

Lambert, J. C. et al. Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).

21. 21.

Pereira, R. M. et al. Protective molecular mechanisms of clusterin against apoptosis in cardiomyocytes. *Heart Fail. Rev.* **23**, 123–129 (2018).

22. 22.

Thangaraj, S. S. et al. Contact activation-induced complex formation between complement factor H and coagulation factor XIIa. *J. Thromb. Haemost.* **18**, 876–884 (2020).

23. 23.

Hunt, L. C., Upadhyay, A., Jazayeri, J. A., Tudor, E. M. & White, J. D. An anti-inflammatory role for leukemia inhibitory factor receptor signaling in regenerating skeletal muscle. *Histochem. Cell Biol.* **139**, 13–34 (2013).

24. 24.

Yamagishi, S. et al. Pigment epithelium-derived factor inhibits TNF-alpha-induced interleukin-6 expression in endothelial cells by suppressing NADPH oxidase-mediated reactive oxygen species generation. *J. Mol. Cell. Cardiol.* **37**, 497–506 (2004).

25. 25.

Pohlkamp, T., Wasser, C. R. & Herz, J. Functional roles of the interaction of APP and lipoprotein receptors. *Front. Mol. Neurosci.* **10**, 54 (2017).

26. 26.

Yang, A. C. et al. Physiological blood-brain transport is impaired with age by a shift in transcytosis. *Nature* **583**, 425–430 (2020).

27. 27.

Middeldorp, J. et al. Preclinical assessment of young blood plasma for Alzheimer disease. *JAMA Neurol.* **73**, 1325–1333 (2016).

28. 28.

Zhou, Z., Xu, M. J. & Gao, B. Hepatocytes: a key cell type for innate immunity. *Cell. Mol. Immunol.* **13**, 301–315 (2016).

29. 29.

Yin, C. et al. ApoE attenuates unresolvable inflammation by complex formation with activated C1q. *Nat. Med.* **25**, 496–506 (2019).

30. 30.

Hong, S. et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **352**, 712–716 (2016).

31. 31.

Kirszbaum, L., Bozas, S. E. & Walker, I. D. SP-40,40, a protein involved in the control of the complement pathway, possesses a unique array of disulphide bridges. *FEBS Lett.* **297**, 70–76 (1992).

32. 32.

Hsu, J. L. et al. Plasma biomarkers are associated with agitation and regional brain atrophy in Alzheimer's disease. *Sci. Rep.* **7**, 5035 (2017).

33. 33.

Yousef, H. et al. Aged blood impairs hippocampal neural precursor activity and activates microglia via brain endothelial cell VCAM1. *Nat. Med.* **25**, 988–1000 (2019).

34. 34.

Lee, J. W. et al. Neuro-inflammation induced by lipopolysaccharide causes cognitive impairment through enhancement of beta-amyloid generation. *J. Neuroinflammation* **5**, 37 (2008).

35. 35.

Monje, M. L., Toda, H. & Palmer, T. D. Inflammatory blockade restores adult hippocampal neurogenesis. *Science* **302**, 1760–1765 (2003).

36. 36.

Rockenstein, E., Mallory, M., Mante, M., Sisk, A. & Masliah, E. Early formation of mature amyloid- $\beta$  protein deposits in a mutant APP transgenic model depends on levels of A $\beta$ <sub>1–42</sub>. *J. Neurosci. Res.* **66**, 573–582 (2001).

37. 37.

Villeda, S. A. et al. The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature* **477**, 90–94 (2011).

38. 38.

Liddelow, S. A. et al. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).

39. 39.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

40. 40.

Tabula Muris, C. et al. Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018).

41. 41.

Hahn, O. et al. A nutritional memory effect counteracts benefits of dietary restriction in old mice. *Nat. Metab.* **1**, 1059–1073 (2019).

42. 42.

Brake, M. A. et al. Assessing blood clotting and coagulation factors in mice. *Curr. Protoc. Mouse Biol.* **9**, e61 (2019).

43. 43.

Zheng, Z. et al. An ATF6-tPA pathway in hepatocytes contributes to systemic fibrinolysis and is repressed by DACH1. *Blood* **133**, 743–753 (2019).

44. 44.

Smith, A. A., Jacobson, L. J., Miller, B. I., Hathaway, W. E. & Manco-Johnson, M. J. A new euglobulin clot lysis assay for global

fibrinolysis. *Thromb. Res.* **112**, 329–337 (2003).

45. 45.

Yousef, H., Czupalla, C. J., Lee, D., Butcher, E. C. & Wyss-Coray, T. Papain-based single cell isolation of primary murine brain endothelial cells using flow cytometry. *Bio Protoc.* **8**, e3091 (2018).

46. 46.

Han, Y., Li, M., Qiu, F., Zhang, M. & Zhang, Y. H. Cell-permeable organic fluorescent probes for live-cell long-term super-resolution imaging reveal lysosome-mitochondrion interactions. *Nat. Commun.* **8**, 1307 (2017).

47. 47.

Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).

48. 48.

Hughes, C. S. et al. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).

49. 49.

Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

50. 50.

Rahnenfuhrer A, A. topGO: enrichment analysis for Gene Ontology. R package version 2.18.0 <https://doi.org/10.18129/B9.bioc.topGO> (2016).

51. 51.

Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* **6**, e21800 (2011).

52. 52.

Albert, M. S. et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 270–279 (2011).

53. 53.

Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).

54. 54.

*SOMAscan* Technical White Paper (Somalogic, 2015);  
<http://www.somalogic.com/wp-content/uploads/2016/08/SSM-002-Rev-3-SOMAscan-Technical-White-Paper.pdf>

55. 55.

Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

56. 56.

Schaum, N., et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).

57. 57.

Bostrom, P. et al. A PGC1- $\alpha$ -dependent myokine that drives brown-fat-like development of white fat and thermogenesis. *Nature* **481**, 463–468

(2012).

58. 58.

Cho, H. C. et al. The concentrations of serum, plasma and platelet BDNF are all increased by treadmill VO<sub>2max</sub> performance in healthy college men. *Neurosci. Lett.* **519**, 78–83 (2012).

59. 59.

Colt, E. W., Wardlaw, S. L. & Frantz, A. G. The effect of running on plasma β-endorphin. *Life Sci.* **28**, 1637–1640 (1981).

60. 60.

El Hayek, L. et al. Lactate mediates the effects of exercise on learning and memory through sirt1-dependent activation of hippocampal brain-derived neurotrophic factor (BDNF). *J. Neurosci.* **39**, 2369–2382 (2019).

61. 61.

Eliakim, A. et al. Physical fitness, endurance training, and the growth hormone-insulin-like growth factor I system in adolescent females. *J. Clin. Endocrinol. Metab.* **81**, 3986–3992 (1996).

62. 62.

Eliakim, A., Brasel, J. A., Mohan, S., Wong, W. L. & Cooper, D. M. Increased physical activity and the growth hormone-IGF-I axis in adolescent males. *Am. J. Physiol.* **275**, R308–R314 (1998).

63. 63.

Ferreira, J. C. et al. Maximal lactate steady state in running mice: effect of exercise training. *Clin. Exp. Pharmacol. Physiol.* **34**, 760–765 (2007).

64. 64.

Koehl, M. et al. Exercise-induced promotion of hippocampal cell proliferation requires  $\beta$ -endorphin. *FASEB J.* **22**, 2253–2262 (2008).

65. 65.

Koziris, L. P. et al. Serum levels of total and free IGF-I and IGFBP-3 are increased and maintained in long-term training. *J. Appl. Physiol.* **86**, 1436–1442 (1999).

66. 66.

Moon, H. Y. et al. Running-induced systemic cathepsin B secretion is associated with memory function. *Cell Metab.* **24**, 332–340 (2016).

67. 67.

Schabitz, W. R. et al. Intravenous brain-derived neurotrophic factor reduces infarct size and counterregulates Bax and Bcl-2 expression after temporary focal cerebral ischemia. *Stroke* **31**, 2212–2217 (2000).

68. 68.

Wrann, C. D. et al. Exercise induces hippocampal BDNF through a PGC-1 $\alpha$ /FNDC5 pathway. *Cell Metab.* **18**, 649–659 (2013).

## Acknowledgements

We thank members of the Wyss-Coray laboratory for their support; D. Berdnik for technical assistance; H. du Bois and S. Shuken for sharing their protocols; H. Zhang and K. Dickey for laboratory management; and members of the Stanford Behavioral and Functional Neuroscience Laboratory for their work on the behavioural assays. This work was funded by the US National Institute on Aging (AG047820 to T.W.-C. and T.A.R. and 1F32AG067652 to N.K.) and the Stanford Alzheimer's Disease Research Center (P30AG066515), the US Department of Veterans Affairs (Research Career Scientist Award IO1 BX001319 to T.W.-C.), a NOMIS Foundation award to T.W.-C., the Simons Foundation (award to T.W.-C.), the Wu Tsai Neurosciences Institutes' Brain Rejuvenation Project with

support from the Bertarelli Foundation (award to T.W.-C.), the Department of Defense (W81XWH-12-1-0584 to K.J.F.), the Alzheimer's Association (NIRG-15-362171 to J.K.F.) and the Marie Curie Foundation n-273487 (GCs-CNS-IS) awarded to Z.D.M. The contents supported by this funding do not represent the views of the VA or the US Government. We acknowledge Servier Medical Art (<https://smart.servier.com>), MediaLab (<https://medialab.biochem.wisc.edu/clip-art/>) and Freepik (<https://www.freepik.com>) for providing images of mice and cartoon components.

## Author information

### Author notes

1. Zurine De Miguel

Present address: Psychology Department, California State University, Monterey Bay, CA, USA

2. Benoit Lehallier

Present address: Alkahest Inc, San Carlos, CA, USA

3. Niclas Olsson

Present address: Calico Life Sciences, South San Francisco, CA, USA

4. These authors contributed equally: Nathalie Khoury, Michael J. Betley

## Affiliations

1. Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA

Zurine De Miguel, Nathalie Khoury, Michael J. Betley, Benoit Lehallier, Drew Willoughby, Andrew C. Yang, Oliver Hahn, Nannan

Lu, Ryan T. Vest, Liana N. Bonanno, Davis Lee, Hui Zhang, Thomas A. Rando & Tony Wyss-Coray

2. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA

Zurine De Miguel, Nathalie Khoury, Benoit Lehallier, Drew Willoughby, Andrew C. Yang, Oliver Hahn, Nannan Lu, Ryan T. Vest, Liana N. Bonanno, Davis Lee, Hui Zhang, Thomas A. Rando & Tony Wyss-Coray

3. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA

Zurine De Miguel, Nathalie Khoury, Benoit Lehallier, Drew Willoughby, Andrew C. Yang, Oliver Hahn, Nannan Lu, Ryan T. Vest, Liana N. Bonanno, Davis Lee, Hui Zhang & Tony Wyss-Coray

4. Neurosciences Graduate Training Program, Stanford University School of Medicine, Stanford, CA, USA

Michael J. Betley

5. Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

Niclas Olsson & Joshua E. Elias

6. The Veterans Affairs Palo Alto HealthCare System, Palo Alto, CA, USA

Lakshmi Yerra, J. Kaci Fairchild & Thomas A. Rando

7. Chan Zuckerberg Biohub, Stanford, CA, USA

Lichao Zhang & Joshua E. Elias

8. Behavioral and Functional Neuroscience Laboratory, Stanford University School of Medicine, Stanford, CA, USA

Nay Lui Saw & Mehrdad Shamloo

9. Otolaryngology Head and Neck Surgery Research Division, Stanford University, Stanford, CA, USA

Patrick L. McAlpine

10. Department of Genetics, Stanford University, Stanford, CA, USA

Kévin Contrepois

## Contributions

T.W.-C., T.A.R., Z.D.M. and M.B. designed and conceived the experiments. Z.D.M. and M.B. developed an initial paradigm of plasma transfer from runner to non-runner mice and studied the effect of plasma on neural stem cell activity. Z.D.M. and D.W. performed experiments to generate plasma pools, carried out animal treatments and processed brain tissue and plasma samples for molecular and protein analyses. D.W. performed and analysed behavioural experiments under the supervision of Z.D.M.; L.B., N.K., O.H. and Z.D.M. performed and analysed sequencing experiments. Z.D.M., N.L.S. and M.S. designed and performed behavioural experiments. N.K., A.Y., R.V., N.L. and Z.D.M. designed and performed experiments on BECs. N.K. performed statistical analysis and visualization of the single-cell datasets. H.Z. generated and provided the APP mice. D.L. performed retro-orbital injections of rCLU. L.Y., N.K. and Z.D.M. designed and performed experiments to assess the complement and coagulation cascades. B.L. performed statistical analyses and visualization of protein and gene datasets. J.K.F. performed experiments with humans and collected plasma samples. N.O., J.E.E., L.Z., P.L.M. and K.C. carried out MS analyses. Z.D.M., D.W. and T.W.-C. wrote the manuscript with input from T.A.R.; T.W.-C. and Z.D.M. supervised the study.

## Corresponding author

Correspondence to [Tony Wyss-Coray](#).

# Ethics declarations

## Competing interests

N.O. is affiliated with Calico Life Sciences and has no financial interests to declare. The other authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Changes in cell proliferation and survival with running or runner plasma infusions.](#)

**a**, Male mice at 3-4 months of age had access to a running wheel for 3, 7, 14 or 28 days while controls remain without access to running wheels. EdU was administered 24 h before sacrifice. Plots show total number of cells per dentate gyrus (DG) of fluorescent immunolabelled EdU<sup>+</sup> cells (n = 6-8 per group) and DCX<sup>+</sup> cells (n = 6-7 per group). **b**, Running distance per day by male mice (n = 5) at 3-4 months of age with free access to a running wheel. **c**, Male mice at 3, 6, 9, 12 and 15 months of age had access to a running wheel for 28 days. Graphs show total number of DCX<sup>+</sup> cells per dentate gyrus (n=4-8 per group). Means ± s.e.m; unpaired Student's two-tailed *t* test; \* *P* < 0.05, \*\* *P* < 0.01, \*\*\* *P* < 0.001. **d**, Male mice at 3 months of age had access to a running wheel for 28 days. BrdU was administered 3 days before the exercise and EdU 24 h before sacrifice. Graphs show total number of cells per dentate gyrus (DG) of fluorescent immunolabelled cells (n=4-6 per group). Means ± s.e.m; unpaired Student's two-tailed *t* test; \* *P* < 0.05. **e**, Control injections of saline via retro-orbital vein in combination

with Isoflurane do not impair neural stem activity. Mice at 3-4 months of age were injected with saline via the tail vein or the retro orbital vein with 200 µl of saline, every 3 days for 28 days. BrdU was administered 3 days before saline administration and EdU 24 h before sacrifice. The hippocampus was dissected and processed for immunohistochemistry. Graphs show total number of cells per dentate gyrus (DG) of fluorescent immunolabelled cells (n=7-8 per group). Means ± s.e.m; Unpaired Student's two-tailed *t* test. **f**, Runner plasma infusions from 28 days runners upregulate proliferation and survival of hippocampal new born cells. Plasma from running mice (3-4 months of age) that run for 7, 14 or 28 days was collected and transferred to matched aged non-running mice, once every 3 days 28 days. BrdU was administered 3 days before plasma administration and EdU 24 h before sacrifice. Graphs show fold change of total number of cells per dentate gyrus (DG) of fluorescent immunolabelled EdU<sup>+</sup> cells, BrdU<sup>+</sup> cells and DCX<sup>+</sup> cells (n = 6-9 per group). Means ± s.e.m; unpaired Student's two-tailed *t* test; \*  $P < 0.05$ , \*\*  $P < 0.01$  and \*\*\*  $P < 0.001$ . The images in **a**, **c**, **d** and **e** were generated using Servier Medical Art (<https://smart.servier.com>) and MediaLab (<https://medialab.biochem.wisc.edu/clip-art/>).

## **Extended Data Fig. 2 Changes in behavioural and hippocampal biological processes in response to runner plasma infusions.**

**a**, Plasma from running mice (3-4 months of age) was collected and transferred to matched aged non-running mice, once every 3 days for 28 days. Mice were then tested for cued memory on the fear conditioning test. Graphs show percentage of freezing behaviour in response to the light/tone cues associated with the fear stimulus in CP and RP recipient mice (n = 18-20 per group). Means ± s.e.m; Unpaired Student's two-tailed *t* test, N.S. not significant. **b**, CP and RP infused mice show comparable activity and velocity in the activity chamber (n = 13-14 per group) and percentage of time spent in the light arena (n = 12 per group). Means ± s.e.m; unpaired Student's two-tailed *t* test; N.S., not significant. **c**, Graphs show distance traveled, thigmotaxis and velocity displayed by mice infused with RP or CP in the water maze test (n = 12 per group). **d**, Hierarchical networks of the abundance of gene ontology (GO) terms (Fisher's exact test,  $P < 0.05$ ) related to biological processes using REVIGO (Resnik measurement, 0.7

distance). GO terms correspond to the DEGs with treatment of CP versus RP shown in Fig. [1h](#).

**Extended Data Fig. 3 Validation of changes in gene expression in response to runner plasma infusions and validation of changes in proteins with running.**

**a**, PCA analysis of common DEGs (Wald test,  $P < 0.05$ ) induced by LPS (SAL-SAL (blue) vs. SAL-LPS (red) and by RP treatment (LPS-CP (grey) vs. LPS-RP (green)) ( $n = 7\text{-}8$  per group). **b**, Graphs show fold changes of relative gene expression of indicated genes measured by qPCR ( $n = 7\text{-}8$  per group). Means  $\pm$  s.e.m; One-way ANOVA and Bonferroni post-hoc; \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  and \*\*\*\*  $P < 0.0001$ . **c**, Validation of proteins captured with shotgun-LC MS1 using TMT-LC MS3 detection and analysis. Heat map depicting the relative levels of the top differentially expressed plasma proteins detected. Unpaired Student's two-tailed  $t$  test; \*  $P < 0.05$ , \*\*  $P < 0.01$  and \*\*\*  $P < 0.001$ . **d**, Classification by PCA analysis using plasma proteins detected by Shotgun-LC-MS1 and significantly changed with running (unpaired Student's two-tailed  $t$  test,  $P < 0.05$ ). Data are Log<sub>2</sub> transformed and missing values imputed by using the mean of each group. Control plasma (grey); runner plasma (orange). ( $n = 8$  per group). **e**, Hierarchical networks of the abundance of gene ontology (GO) terms (Fisher's exact test,  $P < 0.05$ ) related to biological processes using REVIGO (Resnik measurement, 0.7 distance). GO terms correspond to the proteins significantly change with running when comparing CP with RP (unpaired Student's two-tailed  $t$  test,  $P < 0.05$ ) and shown in Fig. [3b](#).

**Extended Data Fig. 4 Assessment of fibrinolysis and measurement of clot formation in runner plasma.**

**a**, The formation and lysis of clots in control and runner's plasma as measured by the Euglobulin Clot Lysis Time [ECLT] assay. Clot formation and lysis is measured under continuous spectrophotometric absorbance readings at 405nm. (Maximum clot formation is defined as the maximum absorbance; Clot lysis time is defined as the time at which the curve reaches an absorbance of 0.05 or less; Time to half lysis is defined as the

time at which the curve reaches 50% of clot lysis) (n=6-7 per group). Means  $\pm$  s.e.m; unpaired Student's two-tailed *t* test; n.s., not significant, \* *P* < 0.05. **b**, Amount of blood lost in runner and control mice after tail clipping as measured by the haemoglobin concentration (n = 7 per group). Means  $\pm$  s.e.m; unpaired Student's two-tailed *t* test; n.s., not significant. **c**, Measuring the activity of the classical, alternative, and lectin complement pathways in control versus runner's plasma. Plots show absorbance at 450nm indicative of levels of membrane C5b-9 attack complex formation induced by IgM, LPS, or Mannose Binding Lectin (MBL) respectively (n = 10–13 per group). Means  $\pm$  s.e.m; Unpaired Student's two-tailed *t* test; n.s., not significant.

### [Extended Data Fig. 5 Changes in plasma proteins in female and male mice in response to running .](#)

**a - b.** Volcano plot showing proteins significantly changed (*P* < 0.05) in RP versus CP in female or male mice. Downregulated proteins (blue); upregulated proteins (red). Unpaired Student's two-tailed *t* test without FDR correction. **c – d.** Representative plasma proteins of the complement and coagulation pathways in male or female mice with 28 days of running (n = 8 per group). Means  $\pm$  s.e.m, Unpaired Student's two-tailed *t* test.

### [Extended Data Fig. 6 Immunodepletion of clusterin in runner plasma abrogates its anti-inflammatory properties on the hippocampus.](#)

**a.** Male mice (3-4 months of age) were injected with LPS and treated with saline (LPS – SAL, n=10), runner plasma (LPS – RP, n=10), runner plasma without CLU (LPS – RP – CLU, n=8), runner plasma without FH (LPS – RP – FH, n=8), runner plasma without LIFR (LPS – RP – LIFR, n=8)) or runner plasma without PDEF (LPS – RP – PDEF, n=8). Heat map representing the relative differences between the mean of each group (zscore of means) on selected inflammatory gene markers in the hippocampus. **b.** Western blotting shows that CLU removal from a plasma sample after depletion via CLU antibody coupled to superparamagnetic beads using a Dynabeads® Antibody Coupling Kit bind to CLU.

**Extended Data Fig. 7 Expression of *Clu* and its receptor *Lrp8* across different cell types and organs of the adult mouse.**

**a.** *Clu* mRNA levels across organs quantified from the bulk RNA-seq dataset conducted on different organs of the adult mouse<sup>56</sup>. **b.** Bar chart showing the top 20 *Clu* expressing cells at the mRNA level among >100 different cell types of the adult mouse quantified from the *Tabula Muris Atlas* dataset<sup>40</sup> (n=4 male and n = 4 female biological replicates). **c.** The mRNA levels of the *Clu* receptor *LRP8/ApoER2* across organs quantified from the bulk RNA-seq dataset conducted on different organs of the adult mouse<sup>56</sup>. **d.** Bar chart showing the top 20 *Lrp8* expressing cells at the mRNA level among >100 different cell types of the adult mouse quantified from the *Tabula Muris Atlas* dataset<sup>40</sup>.

**Extended Data Fig. 8 Experimental design, cell population consistency, and pathways analysis of the scRNA-seq experiment conducted on hippocampal BEC isolated from LPS and CLU treated mice.**

**a**, Confocal representative images show peripherally injected rCLU tagged with Atto-647N or control PBS containing Atto-647N colocalized with cells of the cerebrovasculature, arteries (SMA), capillaries and veins (Tfrc). Scale bars, 10 μm. **b**, Schematic depicting the experimental paradigm followed for the injections of the three groups. 3- to 4-month-old male mice received Saline only injections (Sal), LPS plus saline treatments (LPS), or LPS plus recombinant CLU (LPS+Clu). BECs were isolated from n = 4-5 per group. **c**, tSNE plots showing the cellular proportions, numbers, and distributions in the three experimental groups (Sal, LPS, LPS+Clu). **d**, tSNE plots show distribution of BECs among arterial, capillary, and venous cells by group. Combined tSNE plot for BECs sorted from 3- to 4-month-old male mice (n = 4-5 mice per group) treated with Saline, LPS, and LPS+CLU. (Cells labelled as BEC are of low quality and were excluded from differential expression analysis.). **e**, Scatter plots show a list of selected genes altered in BECs (arterial, capillary, and venous) by acute inflammation (LPS) and reversed by CLU treatment. Coloured genes represent genes that pass the cutoff fold change of 1.1. Green: Genes

increased by inflammation and reversed by CLU. Blue: Genes reduced by inflammation and reversed by CLU. (Log FC: natural logarithm of fold change). **f-g**, Dotplot showing Gene Ontology (GO) Biological Processes terms for BEC genes (Benjamini–Hochberg adjustment test, FDR < 0.05) that decrease or increase by LPS treatment and are reversed by CLU. Genes were selected based on the cutoff fold change of 1.1. The images in **b** were generated using MediaLab (<https://medialab.biochem.wisc.edu/clip-art/>).

**Extended Data Fig. 9 Experimental design, cell population consistency, and pathways analysis of the scRNA-seq experiment conducted on hippocampal BEC isolated from APP and CLU treated mice.**

**a**, Schematic depicting the experimental paradigm followed for the injections of the four groups. 14-months-old male Wild Type (WT) and APP transgenic mice that received no treatment, 17-months-old male APP mice that received saline only injections (APP+Sal) or repetitive CLU injections (APP+Clu) (BECs were isolated from n = 4-5 per group). **b**, tSNE plots showing the cellular proportions, numbers, and distributions in the four experimental groups (WT, APP, APP+Sal, APP+Clu). **c**, tSNE plots show distribution of BECs among arterial, capillary, and venous cells by group. Combined tSNE plot for BECs sorted from 14-month-old male wild-type and APP transgenic mice and 17-month-old male APP mice treated with Saline or CLU (BECs sorted from n = 4-5 mice per group). (Cells labelled as BEC are of low quality and were excluded from differential expression analysis.). **d**, Scatter plots show a list of selected genes altered in BECs (arterial, capillary, and venous) by chronic (APP) inflammation and reversed by CLU treatment. Coloured genes represent genes that pass the cutoff fold change of 1.1. Green: Genes increased by inflammation and reversed by CLU. Blue: Genes reduced by inflammation and reversed by CLU. (Log FC: natural logarithm of fold change). **e**, Dotplot showing Gene Ontology (GO) Biological Processes terms for BEC genes (Benjamini–Hochberg adjustment test, FDR < 0.05) that decrease or increased in APP mice compared to wild types and are reversed by CLU. Genes were selected based on the cutoff fold change of 1.1. **f**, Venn Diagram showing the number of unique and common genes that are increased by LPS and APP

and reversed by CLU and reduced by LPS and APP and reversed by CLU (Extended Data Fig. 8f,g and 9e). **g**, Heatmap showing the common 20 genes increased with acute (LPS) inflammation and reversed by rCLU treatment (from Extended Data Fig. 8e panel). **h**, Dot plot showing Gene Ontology (GO) Biological Processes terms for BEC genes (Benjamini–Hochberg adjustment test, FDR < 0.05) of the common genes (from panel f, left) that increased with acute (LPS) and chronic (APP) inflammation and are reversed by Clu treatment. The images in **a** were generated using MediaLab (<https://medialab.biochem.wisc.edu/clip-art/>).

### **Extended Data Fig. 10 Changes in cortical gene expression in response to LPS inoculation and treatment with CLU.**

**a.** Number of DEGs (Wald test;  $P < 0.1$ ) between Saline-treated mice ( $n = 4$ ) and either LPS ( $n = 8$ ) or LPS+rCLU ( $n = 7$ ) injected animals. **b.** Venn diagrams depicting the overlap of DEGs (shown in panel a) in whole cortex of mice inoculated with LPS or LPS+rCLU, relative to the Saline-injected controls. **c – d.** Boxplot representation of scaled expression levels of 200 differentially up- and down-regulated genes under LPS as opposed to Saline controls. Whiskers represent the first and fifth quartiles, box edges represent the second and fourth quartiles and the centre line represents the third quartile/median. Two-sided Wilcoxon rank-sum test. **e – f.** Representative GO enrichment categories of LPS significantly regulated genes. Lengths of bars represent negative ln-transformed  $P$  using two-sided Fisher's exact test. Colours indicate gene-wise log2 fold-changes (log2(FC)) between (e) LPS and Saline treatment or (f) LPS+rCLU and LPS treatment. Numbers beside bars indicate DEGs induced by LPS in that GO category.

### **Extended Data Fig. 11 Changes in plasma proteins of the complement and coagulation pathways in humans and male mice in response to exercise.**

Schematic representation of the significantly changed ( $P < 0.05$ ) plasma proteins of complement and coagulation cascades in humans and mice after exercise. Dotted black arrows indicate relationship between factors of the

complement and the coagulation system. The diagram was generated using Servier Medical Art (<https://smart.servier.com>), MediaLab (<https://medialab.biochem.wisc.edu/clip-art/>) and Freepik (<https://www.freepik.com>).

## Extended Data Table 1 Exercise-induced factors that affect the brain

# Supplementary information

## Supplementary Information

A guide to Supplementary Tables 1–20.

## Reporting Summary

## Supplementary Tables

This file contains Supplementary Tables 1–20 – see Supplementary Information document for full descriptions..

# Source data

## Source Data Fig. 1

## Source Data Fig. 3

## Source Data Fig. 4

# Rights and permissions

## Reprints and Permissions

# About this article

## Cite this article

De Miguel, Z., Khoury, N., Betley, M.J. *et al.* Exercise plasma boosts memory and dampens brain inflammation via clusterin. *Nature* **600**, 494–499 (2021). <https://doi.org/10.1038/s41586-021-04183-x>

- Received: 26 July 2019
- Accepted: 26 October 2021
- Published: 08 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04183-x>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04183-x>

- Article
- [Published: 08 December 2021](#)

# Combinatorial, additive and dose-dependent drug–microbiome associations

- [Sofia K. Forslund](#) ORCID: [orcid.org/0000-0003-4285-6993<sup>1,2,3,4,5,6 na1</sup>](https://orcid.org/0000-0003-4285-6993),
- [Rima Chakaroun](#)<sup>7 na1</sup>,
- [Maria Zimmermann-Kogadeeva](#) ORCID: [orcid.org/0000-0001-6021-1246<sup>1 na1</sup>](https://orcid.org/0000-0001-6021-1246),
- [Lajos Markó](#) ORCID: [orcid.org/0000-0003-1888-5575<sup>2,4,5 na1</sup>](https://orcid.org/0000-0003-1888-5575),
- [Judith Aron-Wisnewsky](#)<sup>8,9 na1</sup>,
- [Trine Nielsen](#) ORCID: [orcid.org/0000-0002-2066-7895<sup>10 na1</sup>](https://orcid.org/0000-0002-2066-7895),
- [Lucas Moitinho-Silva](#)<sup>1</sup>,
- [Thomas S. B. Schmidt](#)<sup>1</sup>,
- [Gwen Falony](#) ORCID: [orcid.org/0000-0003-2450-0782<sup>11</sup>](https://orcid.org/0000-0003-2450-0782),
- [Sara Vieira-Silva](#) ORCID: [orcid.org/0000-0002-4616-7602<sup>11</sup>](https://orcid.org/0000-0002-4616-7602),
- [Solia Adriouch](#)<sup>8</sup>,
- [Renato J. Alves](#) ORCID: [orcid.org/0000-0002-7212-0234<sup>1</sup>](https://orcid.org/0000-0002-7212-0234),
- [Karen Assmann](#)<sup>8</sup>,
- [Jean-Philippe Bastard](#)<sup>12,13</sup>,
- [Till Birkner](#) ORCID: [orcid.org/0000-0003-2656-2821<sup>2,3</sup>](https://orcid.org/0000-0003-2656-2821),
- [Robert Caesar](#) ORCID: [orcid.org/0000-0003-1017-2933<sup>14</sup>](https://orcid.org/0000-0003-1017-2933),
- [Julien Chilloux](#) ORCID: [orcid.org/0000-0002-0384-7317<sup>15</sup>](https://orcid.org/0000-0002-0384-7317),
- [Luis Pedro Coelho](#)<sup>1,16,17</sup>,
- [Leopold Fezeu](#)<sup>18</sup>,
- [Nathalie Galleron](#)<sup>19</sup>,
- [Gerard Helft](#)<sup>20</sup>,

- [Richard Isnard<sup>20</sup>](#),
- [Boyang Ji](#) [ORCID: orcid.org/0000-0002-7269-4342<sup>21</sup>](#),
- [Michael Kuhn](#) [ORCID: orcid.org/0000-0002-2841-872X<sup>1</sup>](#),
- [Emmanuelle Le Chatelier<sup>19</sup>](#),
- [Antonis Myridakis](#) [ORCID: orcid.org/0000-0003-1690-6651<sup>15</sup>](#),
- [Lissa Olsson](#) [ORCID: orcid.org/0000-0001-9730-1915<sup>14</sup>](#),
- [Nicolas Pons<sup>19</sup>](#),
- [Edi Prifti](#) [ORCID: orcid.org/0000-0001-8861-1305<sup>8,22,23</sup>](#),
- [Benoit Quinquis<sup>19</sup>](#),
- [Hugo Roume<sup>19</sup>](#),
- [Joe-Elie Salem](#) [ORCID: orcid.org/0000-0002-0331-3307<sup>24</sup>](#),
- [Nataliya Sokolovska<sup>8</sup>](#),
- [Valentina Tremaroli<sup>14</sup>](#),
- [Mireia Valles-Colomer](#) [ORCID: orcid.org/0000-0002-1988-6054<sup>11</sup>](#),
- [Christian Lewinter<sup>25</sup>](#),
- [Nadja B. Søndertoft](#) [ORCID: orcid.org/0000-0001-6350-8117<sup>10</sup>](#),
- [Helle Krogh Pedersen<sup>10</sup>](#),
- [Tue H. Hansen](#) [ORCID: orcid.org/0000-0001-5948-8993<sup>10</sup>](#),
- [The MetaCardis Consortium\\*](#),
- [Jens Peter Gøtze<sup>26</sup>](#),
- [Lars Køber](#) [ORCID: orcid.org/0000-0002-6635-1466<sup>25</sup>](#),
- [Henrik Vestergaard](#) [ORCID: orcid.org/0000-0003-3090-269X<sup>10,27</sup>](#),
- [Torben Hansen](#) [ORCID: orcid.org/0000-0001-8748-3831<sup>10</sup>](#),
- [Jean-Daniel Zucker<sup>8,22,23</sup>](#),
- [Serge Hercberg](#) [ORCID: orcid.org/0000-0002-3168-1350<sup>18</sup>](#),
- [Jean-Michel Oppert<sup>9</sup>](#),
- [Ivica Letunic](#) [ORCID: orcid.org/0000-0003-3560-4288<sup>1,28</sup>](#),
- [Jens Nielsen](#) [ORCID: orcid.org/0000-0002-9955-6003<sup>21</sup>](#),
- [Fredrik Bäckhed](#) [ORCID: orcid.org/0000-0002-4871-8818<sup>10,14,29</sup>](#),
- [S. Dusko Ehrlich](#) [ORCID: orcid.org/0000-0002-7563-4046<sup>19</sup>](#),
- [Marc-Emmanuel Dumas](#) [ORCID: orcid.org/0000-0001-9523-7024<sup>15,30,31,32</sup>](#),
- [Jeroen Raes](#) [ORCID: orcid.org/0000-0002-1337-041X<sup>11</sup>](#),
- [Oluf Pedersen](#) [ORCID: orcid.org/0000-0002-3321-3972<sup>10</sup>](#),

- [Karine Clément](#) ORCID: [orcid.org/0000-0002-2489-3355<sup>8,9</sup>](https://orcid.org/0000-0002-2489-3355),
- [Michael Stumvoll](#) ORCID: [orcid.org/0000-0001-6225-8240<sup>7,33</sup>](https://orcid.org/0000-0001-6225-8240) &
- [Peer Bork](#) ORCID: [orcid.org/0000-0002-2627-833X<sup>1,3,34,35</sup>](https://orcid.org/0000-0002-2627-833X)

*Nature* volume 600, pages 500–505 (2021)

- 4946 Accesses
- 224 Altmetric
- [Metrics details](#)

## Subjects

- [Biomarkers](#)
- [Cardiovascular diseases](#)
- [Systems biology](#)

## Abstract

During the transition from a healthy state to cardiometabolic disease, patients become heavily medicated, which leads to an increasingly aberrant gut microbiome and serum metabolome, and complicates biomarker discovery<sup>1,2,3,4,5</sup>. Here, through integrated multi-omics analyses of 2,173 European residents from the MetaCardis cohort, we show that the explanatory power of drugs for the variability in both host and gut microbiome features exceeds that of disease. We quantify inferred effects of single medications, their combinations as well as additive effects, and show that the latter shift the metabolome and microbiome towards a healthier state, exemplified in synergistic reduction in serum atherogenic lipoproteins by statins combined with aspirin, or enrichment of intestinal *Roseburia* by diuretic agents combined with beta-blockers. Several antibiotics exhibit a quantitative relationship between the number of courses prescribed and progression towards a microbiome state that is associated with the severity of cardiometabolic disease. We also report a relationship between cardiometabolic drug dosage, improvement in clinical markers and microbiome composition, supporting direct drug effects. Taken together,

our computational framework and resulting resources enable the disentanglement of the effects of drugs and disease on host and microbiome features in multimedicated individuals. Furthermore, the robust signatures identified using our framework provide new hypotheses for drug–host–microbiome interactions in cardiometabolic disease.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

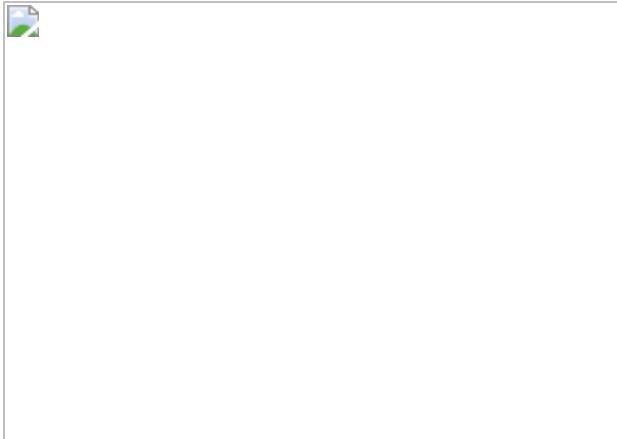
### Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Associations between CMD drugs, host and microbiome.**



**Fig. 2: Combinatorial effects of CMD drugs.**



**Fig. 3: Additive and dose-dependent drug associations with the host and microbiome.**



## Data availability

The source data for the figures are provided at Zenodo (<https://doi.org/10.5281/zenodo.4728981>). Raw shotgun sequencing data that support the findings of this study have been deposited at the ENA under accession codes [PRJEB41311](#), [PRJEB38742](#) and [PRJEB37249](#) with public access. Raw spectra for metabolomics have been deposited in the MassIVE database under the accession codes [MSV000088043](#) (UPLC–MS/MS) and [MSV000088042](#) (GC–MS). The metadata on disease groups and drug intake are provided in Supplementary Tables [1–3](#). The demographic, clinical and phenotype metadata, and processed microbiome and metabolome data for French, German and Danish participants are available at Zenodo (<https://doi.org/10.5281/zenodo.4674360>).

## Code availability

The new drug-aware univariate biomarker testing pipeline is available as an R package (metadeconfoundR; Birkner et al., manuscript in preparation) at Github (<https://github.com/TillBirkner/metadeconfoundR>) and at Zenodo (<https://doi.org/10.5281/zenodo.4721078>). The latest version (0.1.8) of this package was used to generate the data shown in this publication. The code used for multivariate analysis based on the VpThemAll package is available at Zenodo (<https://doi.org/10.5281/zenodo.4719526>). The phenotype and drug intake metadata, processed microbiome, and metabolome data and code resources are available for download at Zenodo (<https://doi.org/10.5281/zenodo.4674360>). The code for reproducing the figures is provided at Zenodo (<https://doi.org/10.5281/zenodo.4728981>).

## References

### 1. 1.

- Sinha, R. et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* **35**, 1077–1086 (2017).

2. 2.

Costea, P. I. et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* **35**, 1069–1076 (2017).

3. 3.

Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).

4. 4.

Schmidt, T. S. B., Raes, J. & Bork, P. The human gut microbiome: from association to modulation. *Cell* **172**, 1198–1215 (2018).

5. 5.

Vujkovic-Cvijin, I. et al. Host variables confound gut microbiota studies of human disease. *Nature* **587**, 448–454 (2020).

6. 6.

Tsuda, A. et al. Influence of proton-pump inhibitors on the luminal microbiota in the gastrointestinal tract. *Clin. Transl. Gastroenterol.* **6**, e89 (2015).

7. 7.

Forslund, K. et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).

8. 8.

Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).

9. 9.

Le Bastard, Q. et al. Systematic review: human gut dysbiosis induced by non-antibiotic prescription medications. *Aliment. Pharmacol. Ther.* **47**, 332–345 (2018).

10. 10.

Vich Vila, A. et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat. Commun.* **11**, 362 (2020).

11. 11.

Vieira-Silva, S. et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* **581**, 310–315 (2020).

12. 12.

Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).

13. 13.

Jackson, M. A. et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat. Commun.* **9**, 2655 (2018).

14. 14.

Conlon, M. A. & Bird, A. R. The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7**, 17–44 (2014).

15. 15.

Blaser, M. J. Antibiotic use and its consequences for the normal microbiome. *Science* **352**, 544–545 (2016).

16. 16.

Imhann, F. et al. Proton pump inhibitors affect the gut microbiome. *Gut* **65**, 740–748 (2016).

17. 17.

Imhann, F. et al. The influence of proton pump inhibitors and other commonly used medication on the gut microbiota. *Gut Microbes* **8**, 351–358 (2017).

18. 18.

Wu, H. et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23**, 850–858 (2017).

19. 19.

Koeth, R. A. et al.  $\gamma$ -Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of l-carnitine to TMAO. *Cell Metab.* **20**, 799–812 (2014).

20. 20.

Kopp, E. & Ghosh, S. Inhibition of NF- $\kappa$ B by sodium salicylate and aspirin. *Science* **265**, 956–959 (1994).

21. 21.

Davalli, A. M., Perego, C. & Folli, F. B. The potential role of glutamate in the current diabetes epidemic. *Acta Diabetol.* **49**, 167–183 (2012).

22. 22.

Schmidt, T. S. et al. Extensive transmission of microbes along the gastrointestinal tract. *eLife* **8**, e42693 (2019).

23. 23.

Shimazu, T. et al. Suppression of oxidative stress by  $\beta$ -hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science* **339**, 211–214 (2012).

24. 24.

Shah, B. R. & Hux, J. E. Quantifying the risk of infectious diseases for people with diabetes. *Diabetes Care* **26**, 510–513 (2003).

25. 25.

Korpela, K. & Vos, W. M. de. Antibiotic use in childhood alters the gut microbiota and predisposes to overweight. *Microbial Cell* **3**, 296–298 (2016).

26. 26.

Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).

27. 27.

Arumugam, M. et al. Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).

28. 28.

Costea, P. I. et al. Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* **3**, 8–16 (2018).

29. 29.

Cameron, A. R. et al. Anti-inflammatory effects of metformin irrespective of diabetes status. *Circ. Res.* **119**, 652–665 (2016).

30. 30.

Dandona, P. et al. Increased plasma concentration of macrophage migration inhibitory factor (MIF) and MIF mRNA in mononuclear

cells in the obese and the suppressive action of metformin. *J. Clin. Endocrinol. Metab.* **89**, 5043–5047 (2004).

31. 31.

Branca, F., Nikogosian, H. & Lobstein, T. The Challenge of Obesity in the WHO European Region and the Strategies for Response. Summary, <https://www.euro.who.int/en/publications/abstracts/challenge-of-obesity-in-the-who-european-region-and-the-strategies-for-response-the-summary> (WHO, 2007).

32. 32.

Alberti, K. G. M. M., Zimmet, P. & Shaw, J. The metabolic syndrome—a new worldwide definition. *Lancet* **366**, 1059–1062 (2005).

33. 33.

Petersmann, A. et al. Definition, classification and diagnosis of diabetes mellitus. *Exp. Clin. Endocrinol. Diabetes* **127**, S1-S7 (2019).

34. 34.

Whelton, P. K. et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Hypertension* **71**, 1269–1324 (2018).

35. 35.

Yancy, C. W. et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines and the Heart Failure Society of America. *J. Card. Fail.* **23**, 628–651 (2017).

36. 36.

Verger, E. O. et al. Dietary assessment in the MetaCardis study: development and relative validity of an online food frequency questionnaire. *J. Acad. Nutr. Diet.* **117**, 878–888 (2017).

37. 37.

Criscuolo, A. & Brisse, S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).

38. 38.

Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).

39. 39.

Cotillard, A. et al. Dietary intervention impact on gut microbial gene richness. *Nature* **500**, 585–588 (2013).

40. 40.

Prifti, E. & Le Chatelier, E. MetaOMineR. A quantitative metagenomics data analyses pipeline v.hal-02800484 (2014).

41. 41.

Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).

42. 42.

Sunagawa, S. et al. Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* **10**, 1196–1199 (2013).

43. 43.

Kultima, J. R. et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* **32**, 2520–2523 (2016).

44. 44.

Coelho, L. P. et al. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. *Microbiome* **7**, 84 (2019).

45. 45.

Prest, E. I., Hammes, F., Kötzsch, S., van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res.* **47**, 7131–7142 (2013).

46. 46.

Vandepitte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).

47. 47.

Katz, M. H. *Multivariable Analysis. A Practical Guide for Clinicians and Public Health Researchers* (Cambridge Univ. Press, 2011).

48. 48.

Costea, P. I. et al. metaSNV: a tool for metagenomic strain level analysis. *PLoS ONE* **12**, e0182392 (2017).

49. 49.

Hahsler, M., Gruen, B., Hornik, K. arules—a computational environment for mining association rules and frequent item sets. *J. Stat. Softw.* **14**, 1–25. (2005).

50. 50.

Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychol. Methods* **15**, 309–334 (2010).

51. 51.

Seabold, S. & Perktold J. statsmodels: econometric and statistical modeling with Python. In *Proc. 9th Python in Science Conference* (2010).

52. 52.

Palleja, A. et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **3**, 1255–1265 (2018).

## Acknowledgements

We thank the MetaCardis participants for their participation in the study, and particularly the patient associations (Alliance du Coeur and CNAO) for their input and interface; D. Bonnefont-Rousselot (Department of Metabolic Biochemistry, Pitié-Salpêtrière Hospital) for the analysis of plasma lipid profiles; and the nurses, technicians, clinical research assistants and data managers from the Clinical Investigation Platform at the Institute of Cardiometabolism and Nutrition for patient investigations, the Clinical Investigation Center (CIC) from Pitié-Salpêtrière Hospital and Human Research Center on Nutrition (CRNH Ile-de-France) as well as the university hospital of Leipzig for the investigation of healthy control individuals. Quanta Medical provided regulatory oversight of the clinical study and contributed to the processing and management of electronic data. This work was supported by the European Union’s Seventh Framework Program for research, technological development and demonstration under grant agreement HEALTH-F4-2012-305312 (METACARDIS). Part of this work was also supported by the EMBL, by the Metagenopolis grant ANR-11-DPBS-0001, by the H2020 European Research Council (ERC-AdG-669830) (to P.B.), and by grants from the Deutsche Forschungsgemeinschaft (SFB1365 to S.K.F. and L.M.; and SFB1052/3 A1 MS to M.S. (209933838)). Assistance Publique-Hôpitaux de Paris is the promoter of the clinical investigation (MetaCardis). M.-E.D. is supported

by the NIHR Imperial Biomedical Research Centre and by grants from the French National Research Agency (ANR-10-LABX-46 (European Genomics Institute for Diabetes)), from the National Center for Precision Diabetic Medicine – PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council (Agreement 20001891/NP0025517) and by the European Metropolis of Lille (MEL, Agreement 2019\_ESR\_11) and by Isite ULNE (R-002-20-TALENT-DUMAS), also jointly funded by ANR (ANR-16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by the European Metropolis of Lille (MEL). R.J.A. is a member of the Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Bioscience. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research institution at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation.

## Author information

### Author notes

1. These authors contributed equally: Sofia K. Forslund, Rima Chakaroun, Maria Zimmermann-Kogadeeva, Lajos Markó, Judith Aron-Wisnewsky, Trine Nielsen

### Affiliations

1. Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

Sofia K. Forslund, Maria Zimmermann-Kogadeeva, Lucas Moitinho-Silva, Thomas S. B. Schmidt, Renato J. Alves, Luis Pedro Coelho, Michael Kuhn, Ivica Letunic & Peer Bork

2. Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany

Sofia K. Forslund, Lajos Markó & Till Birkner

3. Charité–Universitätsmedizin Berlin, Berlin, Germany

Sofia K. Forslund, Till Birkner & Peer Bork

4. Experimental and Clinical Research Center, A Cooperation of Charité–Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany

Sofia K. Forslund & Lajos Markó

5. Berlin Institute of Health (BIH), Berlin, Germany

Sofia K. Forslund & Lajos Markó

6. DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany

Sofia K. Forslund

7. Medical Department III—Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany

Rima Chakaroun, Judith Kammer, Stefanie Walther & Michael Stumvoll

8. Nutrition and Obesities; Systemic Approaches (NutriOmics), Sorbonne Université, INSERM, Paris, France

Judith Aron-Wisnewsky, Solia Adriouch, Karen Assmann, Edi Prifti, Nataliya Sokolovska, Chloe Amouyal, Fabrizio Andreelli, Maria-Carlota Dao, Line Engelbrechtsen, Christine Poitou, Christine Rouault, Timothy Swartz, Jean-Daniel Zucker & Karine Clément

9. Nutrition Department, Assistance Publique Hôpitaux de Paris, Pitie-Salpêtrière Hospital, Paris, France

Judith Aron-Wisnewsky, Fabrizio Andreelli, Dominique Cassuto, Cecile Ciangura, Jean Khemis, Lea Lucas-Martini, Jonathan Medina-Stamminger, Christine Poitou, Camille Vatier, Jean-Michel Oppert & Karine Clément

10. Novo Nordisk Foundation Center for Basic Metabolic Research,  
Faculty of Health and Medical Sciences, University of Copenhagen,  
Copenhagen, Denmark

Trine Nielsen, Nadja B. Søndertoft, Helle Krogh Pedersen, Tue H. Hansen, Ehm Astrid Andersson Galijatovic, Bolette Hartmann, Jens Juul Holst, Malene Hornbak, Johanne Justesen, Nikolaj Krarup, Mathilde Svendstrup, Henrik Vestergaard, Torben Hansen, Fredrik Bäckhed & Oluf Pedersen

11. Center for Microbiology, VIB, and Rega Institute, KU Leuven,  
Leuven, Belgium

Gwen Falony, Sara Vieira-Silva, Mireia Valles-Colomer & Jeroen Raes

12. Assistance Publique Hôpitaux de Paris, UF Biomarqueurs Inflammatoires et Métaboliques, Biochemistry and Hormonology Department, Tenon Hospital, Paris, France

Jean-Philippe Bastard & Soraya Fellahi

13. Centre de Recherche Saint-Antoine, Sorbonne Université, INSERM UMR S938, Paris, France

Jean-Philippe Bastard & Soraya Fellahi

14. Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Sahlgrenska Center for Cardiovascular and Metabolic Research, University of Gothenburg, Gothenburg, Sweden

Robert Caesar, Lisa Olsson, Valentina Tremaroli, Laura Martinez-Gili & Fredrik Bäckhed

15. Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK

Julien Chilloux, Antonis Myridakis, Pilar Galan, Ruby Kozlowski, Ana Luisa Neves, Michael Olanipekun, Andrea Rodriguez-Martinez & Marc-Emmanuel Dumas

16. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

Luis Pedro Coelho

17. MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, and MOE Frontiers Center for Brain Science, Shanghai, China

Luis Pedro Coelho

18. Nutritional Epidemiology Research Team (EREN), Sorbonne Paris Cité Epidemiology and Statistics Research Centre (CRESS), U1153 INSERM, U1125, INRA, CNAM, University of Paris 13, Paris, France

Leopold Fezeu & Serge Hercberg

19. Metagenopolis, INRA, AgroParisTech, Université Paris-Saclay, Paris, France

Nathalie Galleron, Emmanuelle Le Chatelier, Nicolas Pons, Benoit Quinquis, Hugo Roume, Magalie Berland, Hervé Blotti  re, Mickael Camus, Ang  lique Dor  , Sebastien Fromentin, Sophie Jaqueminet, Hanna Julienne, Nicolas Maziers, Laetitia Pasero Le Pavin, Thierry Vanduyvenboden & S. Dusko Ehrlich

20. Cardiology Department, Assistance Publique H  pitaux de Paris, Pitie-Salp  tri  re Hospital, Paris, France

Gerard Helft, Richard Isnard, Olivier Barthelemy, Jean-Paul Batisse, Rachid Boubrit, Jean-Philippe Collet, Morad

Djebbar, Mathieu Kerneis, Véronique Lejard, Florence Levenez, Robin Massey, Gilles Montalescot, Francoise Pousset & Johanne Silvain

21. Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

Boyang Ji & Jens Nielsen

22. Integromics Unit, Institute of Cardiometabolism and Nutrition, Paris, France

Edi Prifti, Eugeni Belda, Timothy Swartz & Jean-Daniel Zucker

23. Sorbonne Université, IRD, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, Paris, France

Edi Prifti, Sandrine Moute & Jean-Daniel Zucker

24. Department of Pharmacology, Sorbonne Université, INSERM CIC-1901, UNICO-GRECO Cardio-oncology ProgramAP-HP.Sorbonne, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

Joe-Elie Salem & Jean-Sebastien Hulot

25. Department of Cardiology, Rigshopitalet, University of Copenhagen, Copenhagen, Denmark

Christian Lewinter & Lars Køber

26. Department of Clinical Biochemistry, Rigshopitalet, University of Copenhagen, Copenhagen, Denmark

Jens Peter Gøtze

27. Department of Medicine, Bornholms Hospital, Rønne, Denmark

Henrik Vestergaard

28. Biobyte Solutions, Heidelberg, Germany

Chloe Amouyal, Fabrizio Andreelli, Frederic Bosquet, Olivier Bourron, Cecile Ciangura, Agnes Hartemann, Niklas Rye Jørgensen & Ivica Letunic

29. Region Västra Götaland, Sahlgrenska University Hospital, Department of Clinical Physiology, Gothenburg, Sweden

Fredrik Bäckhed

30. McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada

Dominique Gauguier & Marc-Emmanuel Dumas

31. Section of Genomic and Environmental Medicine, National Heart & Lung Institute, Faculty of Medicine, Imperial College London, London, UK

Michael Olanipekun & Marc-Emmanuel Dumas

32. European Genomic Institute for Diabetes, CNRS UMR 8199, INSERM UMR 1283, Institut Pasteur de Lille, Lille University Hospital, University of Lille, Lille, France

Marc-Emmanuel Dumas

33. Helmholtz Institute for Metabolic, Obesity and Vascular Research (HIMAG) of the Helmholtz Zentrum München, University of Leipzig, Leipzig, Germany

Michael Stumvoll

34. Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany

Peer Bork

35. Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea

Peer Bork

36. Biochemistry Department of Metabolic Disorders, Assistance Publique Hôpitaux de Paris, Pitie-Salpêtrière Hospital, Paris, France

Randa Bittar

37. UMRS 1124 INSERM, Université de Paris Descartes, Paris, France

Dominique Gauguier

38. Assistance Publique Hôpitaux de Paris, Endocrinology Department, Pitie-Salpêtrière Hospital, Paris, France

Philippe Giral & Laurence Pouzoulet

39. Department of Biosciences, Nottingham Trent University, Nottingham, UK

Lesley Hoyles

40. PARCC, INSERM, Université de Paris, Paris, France

Jean-Sebastien Hulot

41. Assistance Publique Hôpitaux de Paris, Hôpital Européen Georges-Pompidou, CIC1418 and DMU CARTE, Paris, France

Jean-Sebastien Hulot

42. Integrative Phenomics, Paris, France

Timothy Swartz

## **Consortia**

### **The MetaCardis Consortium\***

- Chloe Amouyal
- , Ehm Astrid Andersson Galijatovic
- , Fabrizio Andreelli
- , Olivier Barthelemy
- , Jean-Paul Batisse
- , Eugeni Belda
- , Magalie Berland
- , Randa Bittar
- , Hervé Blotti re
- , Frederic Bosquet
- , Rachid Boubrit
- , Olivier Bourron
- , Mickael Camus
- , Dominique Cassuto
- , Cecile Ciangura
- , Jean-Philippe Collet
- , Maria-Carlota Dao
- , Morad Djebbar
- , Ang lique Dor 
- , Line Engelbrechtsen
- , Soraya Fellahi
- , Sebastien Fromentin
- , Pilar Galan
- , Dominique Gauguier
- , Philippe Giral
- , Agnes Hartemann
- , Bolette Hartmann
- , Jens Juul Holst
- , Malene Hornbak
- , Lesley Hoyles
- , Jean-Sebastien Hulot
- , Sophie Jaqueminet
- , Niklas Rye J rgensen
- , Hanna Julienne
- , Johanne Justesen
- , Judith Kammer
- , Nikolaj Krarup

- , Mathieu Kerneis
- , Jean Khemis
- , Ruby Kozlowski
- , Véronique Lejard
- , Florence Levenez
- , Lea Lucas-Martini
- , Robin Massey
- , Laura Martinez-Gili
- , Nicolas Maziers
- , Jonathan Medina-Stamminger
- , Gilles Montalescot
- , Sandrine Moute
- , Ana Luisa Neves
- , Michael Olanipekun
- , Laetitia Pasero Le Pavin
- , Christine Poitou
- , Francoise Pousset
- , Laurence Pouzoulet
- , Andrea Rodriguez-Martinez
- , Christine Rouault
- , Johanne Silvain
- , Mathilde Svendstrup
- , Timothy Swartz
- , Thierry Vanduyvenboden
- , Camille Vatier
- & Stefanie Walther

## **Contributions**

K.C. (coordinator), P.B., M.S., O.P., S.D.E., J.R., M.-E.D., F.B. and J.N. conceived the overall objectives and study design of the MetaCardis initiative. S.K.F. and P.B. developed the present project concept and protocol and supervised the project. MetaCardis cohort recruitment, phenotyping and lifestyle recording were conducted by R.C., J.A.-W., T.N., C.L., L.K., T.H., T.H.H., H.V. and K.A., and supervised by M.S., K.C. and O.P. Data curation was undertaken by S.K.F., R.C., L.M., K.A., J.A.-W. and T.N.. Faecal microbial DNA

extraction and shotgun sequencing was performed by N.P., E.L.C. and S.F. Bacterial cell count measurement was performed by G.F. and S.V.-S. Serum and urine metabolome profiling was performed by L.H., J.C., A.M. and M.O. Bioinformatics and statistical analyses were performed by T.B., M.Z.-K., S.K.F., L.S., T.S.B.S., L.P.C., N.S., J.Z., E.P., S.F., R.C., S.V., G.F. and B.J. The manuscript was drafted by S.K.F., R.C., M.Z.-K. and L.M. All of the authors participated in the project development, revision of Article and approved the final version for publication.

## Corresponding authors

Correspondence to [Karine Clément](#), [Michael Stumvoll](#) or [Peer Bork](#).

## Ethics declarations

## Competing interests

F.B. is shareholder in Implexion Pharma AB. K.C. is a consultant for Danone Research, LNC therapeutics and CONFO therapeutics for work that is unassociated with the present study. K.C. has held a collaborative research contract with Danone Research in the context of MetaCardis project. M.B. received lecture and/or consultancy fees from AstraZeneca, Boehringer-Ingelheim, Lilly, Novo Nordisk, Novartis and Sanofi. The other authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Peter Turnbaugh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

## Extended Data Fig. 1 A post-hoc testing approach for deconfounding univariate biomarker analysis for multiple medications and risk factors.

The schematic highlights our covariate control approach. All significant associations between putative drivers (e.g., disease D) and covariates ( $C_1 \dots C_n$ ) to each measured feature ( $Y_1 \dots Y_m$ ) are taken. The outcome of the test is denoted with  $a_i$  for a positive outcome (“yes”) and  $\bar{a}_i$  for a negative outcome (“no”). A significant predictor is called “confounded” and is filtered out in a post-hoc test if there is at least one covariate (e.g., drug treatment or combination) such that the predictor does not add significant predictive capacity beyond the covariate (“confounded”). If no such covariate itself passes the same test (i.e., covariates cannot in turn be shown to have predictive capacity beyond tested predictor), the predictor is considered ambiguous (“ambiguously deconfounded”). Otherwise, the predictor is considered “confidently deconfounded” (we note that “confidently deconfounded” is defined as no confounders were found among all covariates measured in our study).

## Extended Data Fig. 2 Previously reported metabolic disease associations are replicated in the MetaCardis cohort under drug deconfounding, highlighting systemic inflammation, short-chain fatty acid and branched-chain amino acid mechanisms underlying insulin resistance.

Cuneiform plot marker hues and direction show sign of effect size (Cliff's delta), intensity and size show amplitude of effect size, comparing metabolic diseased proband subsets (horizontal axis) with healthy control subject in the MetaCardis population for different microbiome, metabolome and host features (vertical axis). Bold and opaque markers show significant associations (two-sided MWU FDR  $< 0.1$ ) not reducible to any significant drug or demographic confounder. Full associations are found in Supplementary Table 9; here a preselected subset is displayed reflecting previously reported risk and protective factors, validated in MetaCardis.  $^1H$  NMR features are shown with retention time in parentheses, functional

modules with GMM or KEGG identifier in parenthesis, analogous for metagenomic species and mOTUs.

**Extended Data Fig. 3 Previously reported drug-microbiome associations are replicated in the MetaCardis cohort for metformin and PPI.**

Bar plots show the magnitude and direction of effect size (Cliff's delta) of metformin treatment (left) and PPI treatment (right) on microbiome features. These effects are compared to the previously published data from two independent patient cohorts<sup>10</sup>. Only features with direct match on the taxonomic level were included in the comparison. Full list of associations is provided in Supplementary Table 6.

**Extended Data Fig. 4 SNV analysis of strains in the gut of subjects taking PPIs.**

Cuneiform plot shows change in abundance of bacterial species in the gut in subjects taking/not taking PPIs (controlling for other drugs and demographic factors) in each clinical group separately, and for all subjects pooled together. Rows marked “SNV” show whether oral strain single nucleotide markers are significantly (two-sided MWU FDR < 0.1) enriched over gut strain markers in subjects taking PPIs, controlling for abundance of each species. Marker direction, colour and size denote the sign and value of Cliff's delta standardized effect size; opaque markers are significantly altered (two-sided MWU FDR < 0.1; passing all confounder checks). Bacteria are shown if their abundance is significantly altered under PPI consumption, and there are SNPs distinguishing oral from gut strains in HMP samples. (See Supplementary Tables 5–7).

**Extended Data Fig. 5 Breakdown of antibiotics association into individual features, selected features shown.**

Left cuneiform plot (markers show Spearman correlation direction by shape and colour, scope by size and colour, significance (two-sided MWU FDR < 0.1, deconfounded for other drug and demographic features) by edge

opacity) shows association between each feature and total number of antibiotics courses in CMD groups as well as in healthy controls. Right cuneiform shows whether the same features are significantly different (two-sided MWU FDR < 0.1) between healthy controls and CMD subjects following drug deconfounding (markers show Cliff's delta effect size), requiring significant and deconfounded correlation with number of antibiotic courses demonstrable in at least one proband group and at least one group showing significant and deconfounded alteration compared to healthy controls. Core features include increased carriage of possible disease-associated *Ruminococcus gnavus* and various *Clostridia* species, alongside decreased carriage of commensals such as *Faecalibacterium* species. Full list of associations is provided in Supplementary Table 12.

### Extended Data Fig. 6 Taxonomic changes are validated in a recent intervention cohort.

For bacterial species where an effect on abundance of total antibiotics courses in MetaCardis could be demonstrated (significant at Spearman FDR < 0.1 and deconfounded), and where effect of antibiotic intervention has been tested in a recent antibiotic intervention study<sup>52</sup>, MetaCardis correlation on vertical axis vs intervention log-transformed fold change on horizontal axis are displayed. Separate markers are shown for each MetaCardis patient group within which antibiotic effect can be demonstrated. Bold markers achieve significance (FDR < 0.1) in the intervention study as well. For the majority of taxa overlapping between studies, direction of changes matches, consistent with a causal impact of antibiotics on the microbiota in MetaCardis.

### Extended Data Fig. 7 Enterotype likelihood is altered by antibiotics.

Cuneiform shows normalized regression coefficients of logistic models for each 4-class enterotype as a function of antibiotics courses in the past 5 years, separately for controls and metabolic disease patient groups. All significant (two-sided Wald FDR < 0.1) models show depletion of *Ruminococcus* and *Prevotella* enterotypes, and enrichment for *Bacteroides*

enterotypes; in the case of metabolic disease patients, this is strongest for the low cell count *Bacteroides* 2 enterotype.

### **Extended Data Fig. 8 Illustration of flow cytometry gating strategy. A fixed gating/staining approach was applied<sup>45</sup>.**

Both blank and sample solutions were stained with SYBR Green I. **a.** FL1-A/FL3-A acquisition plot of a blank sample (0.85% w/v physiological solution) with gate boundaries indicated. A threshold value of 2000 was applied on the FL1 channel. **b.** Secondary gating was performed on the FSC-A/SSC-A channels to further discriminate between debris/background and microbial events. **c, d.** FL1-A/FL3-A count acquisition of a faecal sample with secondary gating on FSC-A/SSC-A channels based on blank analyses. Total counts were defined as events registered in the FL1-A/FL3-A gating area excluding debris/background events observed in the FSC-A/SSC-A R1 gate. The flow rate was set at 14 microliters per minute and the acquisition rate did not exceed 10,000 events per second. Each panel reflects the events registered during a 30 s acquisition period. Cell counts were determined in duplicate starting from a single biological sample.

## **Supplementary information**

### **Supplementary Information**

Supplementary Text (including the Supplementary Results) and Supplementary References.

### **Reporting Summary**

### **Supplementary Tables 1–4**

Supplementary Tables 1–4 include information on the cohort characteristics and drug intake with source metadata. Source metadata for the MetaCardis cohort include disease group, gender, age, study centre, BMI ( $\text{kg m}^{-2}$ ), alternative healthy eating index, diet diversity score and dietary approaches to stop hypertension, physical activity and manual work levels, and

smoking status. Source data include information on the drug intake, drug combinations, dosage and antibiotic use analysed in the MetaCardis cohort.

### **Supplementary Table 5**

Multivariate breakdown of variance. The table includes multivariate breakdown of variance for each feature space by each predictor, a summary of all interaction terms in the models and the results of the confounder analysis for all patient groups analysed in the MetaCardis cohort.

### **Supplementary Table 6**

Features of microbiome, host and metabolome impacted by different drug groups and drug compounds. Results of drug group (or drug compound according to the ATC classification) assessment for its impact on host and microbiome features for each patient group. The compound comparison with ref.<sup>8</sup> tab shows microbiome features that are negatively impacted by the drug treatment (for the ATC-level compounds) in at least one patient group, and bacterial species of which the growth was inhibited by the same drug in the in vitro experiment.

### **Supplementary Table 7**

The number of microbiome and metabolome features impacted by different drug groups. Summary statistics of features impacted by different drugs, separated by host and microbiome features, and defined as either drug effects (drug effect discordant with the disease effect) or severity markers (drug effect concordant with the disease effect).

### **Supplementary Table 8**

Features of the microbiome, host and metabolome impacted by different drug combinations. Analysis of the effect of drug combinations, assessed for impact on host and microbiome falling within different measurement categories in each patient group.

## **Supplementary Table 9**

Features of microbiome, host and metabolome impacted by patient group/clinical indication. Analysis of the patient groups, contrasted against healthy control individuals, assessed for impact on host and microbiome features falling within different measurement categories.

## **Supplementary Table 10**

Mediation analysis of host and microbiome features for drug intake, dosage and combinations. Mediation analysis using a regression model of drug effect on each host feature mediated through a microbiome feature or vice versa.

## **Supplementary Table 11**

Driver analysis of antibiotic effects on the gut microbiota. Results of the principal coordinate analysis of microbiome composition (cell-count-adjusted mOTU abundance at the species level) performed on antibiotics-naive individuals with T2D as well as on healthy individuals.

## **Supplementary Table 12**

Features of the microbiome, host and metabolome impacted by the number of antibiotic courses during the study period. The results of additive antibiotic exposure effect assessment for its impact on host and microbiome features for each patient group.

## **Supplementary Table 13**

Features of the microbiome, host and metabolome impacted by different drug dosages. The results of drug dosage assessment for its impact on host and microbiome features falling within different measurement categories for each patient group. Both dosage-confirmed (the effect was identified both from drug intake status and relative drug dosage analysis) and dosage-

unique (the effect was revealed only by relative dosage analysis) effects are shown.

## **Supplementary Table 14**

Significant impacts on enterotype distribution based on disease status and medication variables. The table includes clinical status (patient versus healthy control comparisons), CMD or antibiotic drug status or dosage, or intake of drug combinations, shown for each enterotype versus the other three enterotypes in the four-enterotype classification.

## **Rights and permissions**

### Reprints and Permissions

## **About this article**

### **Cite this article**

Forslund, S.K., Chakaroun, R., Zimmermann-Kogadeeva, M. *et al.* Combinatorial, additive and dose-dependent drug–microbiome associations. *Nature* **600**, 500–505 (2021). <https://doi.org/10.1038/s41586-021-04177-9>

- Received: 07 September 2019
- Accepted: 22 October 2021
- Published: 08 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04177-9>

## **Share this article**

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04177-9>

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 14 October 2021](#)

# Genomic reconstruction of the SARS-CoV-2 epidemic in England

- [Harald S. Vöhringer](#)<sup>1</sup>,
- [Theo Sanderson](#)<sup>2,3</sup>,
- [Matthew Sinnott](#) [ORCID: orcid.org/0000-0002-3054-7846](#)<sup>2</sup>,
- [Nicola De Maio](#)<sup>1</sup>,
- [Thuy Nguyen](#)<sup>2</sup>,
- [Richard Goater](#) [ORCID: orcid.org/0000-0001-9954-841X](#)<sup>2</sup>,
- [Frank Schwach](#)<sup>2,4</sup>,
- [Ian Harrison](#) [ORCID: orcid.org/0000-0003-4117-961X](#)<sup>4</sup>,
- [Joel Hellewell](#) [ORCID: orcid.org/0000-0003-2683-0849](#)<sup>5</sup>,
- [Cristina V. Ariani](#)<sup>2</sup>,
- [Sonia Gonçalves](#)<sup>2</sup>,
- [David K. Jackson](#) [ORCID: orcid.org/0000-0002-8090-9462](#)<sup>2</sup>,
- [Ian Johnston](#)<sup>2</sup>,
- [Alexander W. Jung](#)<sup>1</sup>,
- [Callum Saint](#) [ORCID: orcid.org/0000-0001-8720-9736](#)<sup>2</sup>,
- [John Sillitoe](#)<sup>2</sup>,
- [Maria Suciu](#)<sup>2</sup>,
- [Nick Goldman](#) [ORCID: orcid.org/0000-0001-8486-2211](#)<sup>1</sup>,
- [Jasmina Panovska-Griffiths](#) [ORCID: orcid.org/0000-0002-7720-1121](#)<sup>6</sup>,
- [The Wellcome Sanger Institute COVID-19 Surveillance Team](#),
- [The COVID-19 Genomics UK \(COG-UK\) Consortium\\*](#),
- [Ewan Birney](#) [ORCID: orcid.org/0000-0001-8314-8497](#)<sup>1</sup>,
- [Erik Volz](#) [ORCID: orcid.org/0000-0001-6268-8937](#)<sup>7</sup>,
- [Sebastian Funk](#) [ORCID: orcid.org/0000-0002-2842-3406](#)<sup>5</sup>,
- [Dominic Kwiatkowski](#) [ORCID: orcid.org/0000-0002-5023-0176](#)<sup>2</sup>,
- [Meera Chand](#)<sup>4,8</sup>,
- [Inigo Martincorena](#) [ORCID: orcid.org/0000-0003-1122-4416](#)<sup>2</sup>,
- [Jeffrey C. Barrett](#) [ORCID: orcid.org/0000-0002-1152-370X](#)<sup>2</sup> &

- [Moritz Gerstung](#)    [ORCID: orcid.org/0000-0001-6709-963X<sup>1,9</sup>](#)

[\*Nature\*](#) volume 600, pages 506–511 (2021)

- 16k Accesses
- 1 Citations
- 954 Altmetric
- [Metrics details](#)

## Subjects

- [Epidemiology](#)
- [Evolutionary genetics](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

## Abstract

The evolution of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus leads to new variants that warrant timely epidemiological characterization. Here we use the dense genomic surveillance data generated by the COVID-19 Genomics UK Consortium to reconstruct the dynamics of 71 different lineages in each of 315 English local authorities between September 2020 and June 2021. This analysis reveals a series of subepidemics that peaked in early autumn 2020, followed by a jump in transmissibility of the B.1.1.7/Alpha lineage. The Alpha variant grew when other lineages declined during the second national lockdown and regionally tiered restrictions between November and December 2020. A third more stringent national lockdown suppressed the Alpha variant and eliminated nearly all other lineages in early 2021. Yet a series of variants (most of which contained the spike E484K mutation) defied these trends and persisted at moderately increasing proportions. However, by accounting for sustained introductions, we found that the transmissibility of these variants is unlikely to have exceeded the transmissibility of the Alpha variant. Finally, B.1.617.2/Delta was repeatedly introduced in England and grew rapidly in early summer 2021, constituting approximately 98% of sampled SARS-CoV-2 genomes on 26 June 2021.

[Download PDF](#)

## Main

The SARS-CoV-2 virus accumulates approximately 24 point mutations per year, or 0.3 mutations per viral generation<sup>1,2,3</sup>. Most of these mutations appear to be evolutionarily neutral but, as the SARS-CoV-2 epidemic spread around the world during spring 2020, it became apparent that the virus is continuing to adapt to its human host. An initial sign was the emergence and global spread of the spike protein variant D614G in the second quarter of 2020. Epidemiological analyses estimated that this mutation, which defines the B.1 lineage, confers a 20% transmissibility advantage over the original A lineage that was isolated in Wuhan, China<sup>4</sup>.

A broad range of lineages have been defined since that can be used to track SARS-CoV-2 transmission across the globe<sup>5,6</sup>. For example, B.1.177/EU-1 emerged in Spain in early summer 2020 and spread across Europe through travel<sup>7</sup>. Subsequently, four variants of concern (VOCs) have been identified by the WHO and other public health authorities: the B.1.351/Beta lineage was discovered in South Africa<sup>8</sup>, where it spread rapidly in late 2020. The B.1.1.7/Alpha lineage was first observed in Kent in September 2020 (ref. <sup>9</sup>) from where it swept through the United Kingdom and large parts of the world due to a 50–60% increase<sup>10,11,12,13</sup> in transmissibility. P.1/Gamma originated in Brazil<sup>14,15</sup> and has spread throughout South America. Most recently, B.1.617.2/Delta was associated with a large surge of coronavirus disease 2019 (COVID-19) in India in April 2021 and subsequently around the world.

## Epidemiology of SARS-CoV-2 in England

In the United Kingdom, by late June 2021 the COVID-19 Genomics UK Consortium (COG-UK) had sequenced close to 600,000 viral samples. These data have enabled a detailed reconstruction of the dynamics of the first wave of the epidemic in the United Kingdom between February and August 2020 (ref. <sup>16</sup>). Here we leverage a subset of those data—genomic surveillance data generated at the Wellcome Sanger Institute—to characterize the growth rates and geographical spread of different SARS-CoV-2 lineages and reconstruct how newly emerging variants changed the course of the epidemic.

Our data cover England between 1 September 2020 and 26 June 2021, encompassing three epidemic waves and two national lockdowns (Fig. [1a](#)). In this time period, we sequenced 281,178 viral genomes, corresponding to an average of 7.2% (281,178/3,894,234) of all of the positive tests from PCR testing for the wider population, ranging from 5% in winter 2020 to 38% in early summer 2021, and filtered to remove cases that were associated with international travel ([Methods](#) and Extended Data Fig. [1a, b](#)). Overall, a total of 328 SARS-CoV-2 lineages were identified using the PANGO lineage definition<sup>5</sup>. As some of these lineages were only rarely and intermittently detected, we collapsed these on the basis of the underlying

phylogenetic tree into a set of 71 lineages for modelling (Fig. [1b-d](#) and Supplementary Tables [1](#) and [2](#)).

**Fig. 1: SARS-CoV-2 surveillance sequencing in England between September 2020 and June 2021.**

 figure1

**a**, Positive Pillar 2 SARS-CoV-2 tests in England. **b**, The relative frequency of 328 different PANGO lineages, representing approximately 7.2% of the tests shown in **a**. **c**, Positive tests (row 1) and the frequency of 4 major lineages (rows 2–5) across 315 English lower tier local authorities. **d**, The absolute frequency of sequenced genomes mapped to 71 PANGO lineages. The blue areas in the pie charts are proportional to the fraction of LTLAs in which a given lineage was observed.

### Source data

These data reveal a diversity of lineages in the fall of 2020 followed by sweeps of the Alpha and Delta variants (Fig. [1b](#) and Supplementary Tables [2](#) and [3](#)). Figure [1c](#) shows the geographical distribution of cases and of different lineages, studied at the level of 315 English lower tier local authorities (LTLAs), administrative regions with approximately 100,000–200,000 inhabitants.

## Modelling the dynamics of SARS-CoV-2

We developed a Bayesian statistical model that tracks the fraction of genomes from different lineages in each LTLA in each week and fits the daily total number of positive Pillar 2 tests ([Methods](#) and Extended Data Fig. [2](#)). The multivariate logistic regression model is conceptually similar to previous approaches in its estimation of relative growth rates<sup>[10,11](#)</sup>. It accounts for differences in the epidemiological dynamics between LTLAs, and enables the introduction of new lineages (Fig. [2a–c](#)). Despite the sampling noise in a given week, the fitted proportions recapitulate the observed proportions of genomes as revealed by 35 example LTLAs covering the geography of England (Fig. [2b,c](#) and Supplementary Notes [1](#) and [2](#)). The quality of fit is confirmed by different probabilistic model selection criteria (Extended Data Fig. [3](#)) and also evident at the aggregated regional level (Extended Data Fig. [4](#)).

**Fig. 2: Spatiotemporal model of 71 SARS-CoV-2 lineages in 315 English LTLAs between September 2020 and June 2021.**

---

 **figure2**

---

**a**, The average growth rates for 71 lineages. Data are median  $\pm$  95% CI. **b**, Lineage-specific relative frequency for 35 selected LTLAs, arranged by longitude and latitude to geographically cover England. **c**, Fitted lineage-specific relative frequency for the same LTLAs as in **b**. **d**, Fitted lineage-specific incidence for the same LTLAs as in **b**.

#### Source data

Although the relative growth rate of each lineage is modelled as identical across LTLAs, the local viral proportions change dynamically due to the timing and rate of introduction of different lineages. The model also calculates total and lineage-specific local incidences and time-dependent growth rates and approximate reproduction numbers  $R_t$  by negative binomial spline fitting of the number of daily positive PCR tests ([Methods](#), Fig. 2d and Extended Data Fig. 2c). Together, this enables a quantitative reconstruction of different periods of the epidemic, which we will discuss in chronological order.

## Multiple subepidemics in autumn 2020

Autumn 2020 was characterized by a surge of cases—concentrated in the north of England—that peaked in November, triggering a second national lockdown (Fig. 1a, c). This second wave initially featured B.1 and B.1.1 sublineages, which were slightly more prevalent in the south and north of England, respectively (Fig. 2b, c). Yet, the proportion of B.1.177 and its geographically diverse sublineages steadily increased across LTLAs from around 25% at the beginning of September to 65% at the end of October. This corresponds to a growth rate of between 8% (growth per 5.1 d; 95% confidence interval (CI) = 7–9%) and 12% (95% CI = 11–13%) greater than that of B.1 or B.1.1. The trend of B.1.177 expansion relative to B.1 persisted throughout January (Extended Data Fig. 5a) and involved a number of monophyletic sublineages that arose in the UK, and similar patterns were observed in Denmark<sup>17</sup> (Extended Data Fig. 5b). Such behaviour cannot easily be explained by international travel, which was the major factor in the initial spread of B.1.1 throughout Europe in summer 2020 (ref. 7). However, the underlying biological mechanism is unclear as the characteristic A222V spike variant is not believed to confer a growth advantage<sup>7</sup>.

## The spread of Alpha during restrictions

The subsequent third wave from December 2020 to February 2021 was almost exclusively driven by Alpha/B.1.1.7, as described previously<sup>10,11,18</sup>. The rapid sweep of Alpha was due to an estimated transmissibility advantage of 1.52 compared with B.1.1 (growth per 5.1 d; 95% CI = 1.50–1.55; Fig. 2a), assuming an unchanged generation interval distribution<sup>19</sup>. The growth advantage is thought to stem, at least in part, from spike mutations that facilitate ACE2 receptor binding (N501Y)<sup>20,21</sup> and furin cleavage (P681H)<sup>22</sup>. Alpha grew during a period of restrictions, which proved to be insufficient to contain its spread (Fig. 3a).

**Fig. 3: Growth of B.1.1.7/Alpha and other lineages in relation to lockdown restrictions between November 2020 and March 2021.**

---

 **figure3**

**a**, Maps and dates of national and regional restrictions in England. Second national lockdown: closed hospitality businesses; contacts  $\leq 2$ , outdoors only; open schools; reasonable excuse needed for leaving home<sup>45</sup>. Tier 1: private indoor gatherings of  $\leq 6$  persons. Tier 2: as tier 1 plus restricted hospitality services; gatherings of  $\leq 6$  in public outdoor places. Tier 3: as tier 2 plus most hospitality businesses closed. Tier 4: as tier

3 but single outdoor contact. Third national lockdown: closed schools with the exception of key workers. **b**, Local lineage-specific  $R_t$  values for Alpha and the average  $R_t$  value (growth per 5.1 d) of all of the other lineages in the same periods. **c**,  $R_t$  values from  $n = 315$  LTLA shown in **b**. The box centre horizontal line indicates the median, box limits show the quartiles, the whiskers extend to 1.5× the interquartile range. **d**, Total and lineage-specific incidence (top) and  $R_t$  values (bottom) for six selected LTLAs during the period of restrictions. **e**, Crude lineage-specific fold changes (odds ratios) in Alpha and other lineages across the second (orange) and third national lockdown (red).

### Source data

The second national lockdown from 5 November to 1 December 2020 successfully reduced the total number of cases, but this masked a lineage-specific increase ( $R_t > 1$ ; defined as growth per 5.1 d) in Alpha and a simultaneous decrease in other hitherto dominant lineages ( $R_t < 1$ ) in 78% (246/315) of LTLAs<sup>23</sup> (Fig. 3b,c). This pattern of Alpha-specific growth during lockdown is supported by a model-agnostic analysis of raw case numbers and proportions of Alpha genomes (Fig. 3e).

Three levels of regionally tiered restrictions were introduced in December 2020 (ref. 24) (Fig. 3a). The areas under different tiers of restrictions visibly and quantitatively coincide with the resulting local  $R_t$  values, with greater  $R_t$  values in areas with lower restrictions (Fig. 3a–c). The reopening caused a surge of cases across all tiers with  $R_t > 1$ , which is also evident in selected time series (Fig. 3d). As Alpha cases surged, more areas were placed under tier 3 restrictions, and stricter tier 4 restrictions were introduced. Nevertheless, Alpha continued to grow ( $R_t > 1$ ) in most areas, presumably driven by increased social interaction over Christmas (Fig. 3c).

After the peak of 72,088 daily cases on 29 December 2020 (Fig. 1a), a third national lockdown was announced on 4 January 2021 (Fig. 3a). The lockdown and increasing immunity derived from infection and increasing vaccination<sup>25</sup> led to a sustained contraction of the epidemic to approximately 5,500 daily cases by 8 March, when restrictions began to be lifted by reopening schools (further steps of easing occurred on 12 April and 17 May). In contrast to the second national lockdown 93% (296/315) of LTLAs exhibited a contraction in both Alpha and other lineages (Fig. 3e).

## Elimination of lineages in early 2021

The lineage-specific rates of decline during the third national lockdown and throughout March 2021 resulted in large differences in lineage-specific incidence.

Cases of Alpha contracted nationally from a peak of around 50,000 daily new cases to approximately 2,750 on 1 April 2021 (Fig. 4a). At the same time, B.1.177—the most prevalent lineage in November 2020—fell to less than an estimated 10 cases per day. Moreover, the incidence of most other lineages present in autumn 2020 was well below 1 after April 2021, implying that the majority of them have been eliminated. The number of observed distinct PANGO lineages declined from a peak of 137 to only 22 in the first week of April 2021 (Fig. 4b). Although this may be attributed in part to how PANGO lineages were defined, we note that the period of contraction did not replenish the genetic diversity lost due to the selective sweep by Alpha (Extended Data Fig. 6).

**Fig. 4: Elimination of SARS-CoV-2 lineages during spring 2021.**

 figure4

**a**, Modelled lineage-specific incidence in England. The colours resemble major lineages as indicated and shades thereof indicate the respective sublineages. **b**, The observed number of PANGO lineages per week.

[Source data](#)

## Refractory variants with E484K mutations

Parallel to the elimination of many formerly dominant SARS-CoV-2 lineages, a number of new variants were imported or emerged (Fig. 4a). These include the VOCs B.1.351/Beta and P.1/Gamma, which carry the spike variant N501Y that is also found in B.1.1.7/Alpha and a similar pair of mutations (K417N/T and E484K) that were each shown to reduce the binding affinity of antibodies from vaccine-derived or convalescent sera<sup>20,26,27,28,29</sup>. The ability to escape from previous immunity is consistent with the epidemiology of Beta in South Africa<sup>8</sup> and especially the surge of Gamma in Manaus<sup>15</sup>. The variants B.1.525/Eta, B.1.526/Iota, B.1.1.318 and P.2/Zeta also harbour E484K spike mutations as per their lineage definition, and sublineages of Alpha and A.23.1 that acquired E484K were found in England (Fig. 5a,b).

**Fig. 5: Dynamics of E484K variants and Delta between January and June 2021.**

 figure5

- a**, The observed relative frequency of other lineages (light grey), Alpha/B.1.1.7 (dark grey), E484K variants (orange) and Delta/B.1.617.2 (brown). **b**, The observed and modelled relative frequency of variants in England. **c**, The total and relative lineage-specific incidence in four selected LTLAs. For **b** and **c**, the shaded areas indicate the 95% CIs. **d**, Estimated UK clade numbers (numbers in square parentheses represent minimum and maximum numbers) and sizes. **e**, Crude growth rates (odds ratios) of Delta and Alpha between April and June 2021, as in Fig. 3e. **f**, Lineage-specific  $R_t$  values of  $n = 315$  LTLA in the same period, defined as in Fig. 3c. **g**, Changes in the average transmissibility across 315 LTLAs during the study period.

#### Source data

The proportion of these E484K-containing variants was consistently 0.3–0.4% from January to early April 2021. A transient rise, especially of the Beta and Gamma variants, was observed in May 2021 (Fig. 5a, b). Yet, the dynamics were largely stochastic and characterized by a series of individual and localized outbreaks, possibly curtailed by local surge testing efforts against Beta and Gamma variants (Fig. 5c). Consistent with the transient nature of these outbreaks, the estimated growth rates of these variants were typically lower than Alpha (Fig. 2a).

Sustained imports from international travel were a critical driving mechanism behind the observed number of non-Alpha cases. A phylogeographical analysis establishing the most parsimonious sets of monophyletic and exclusively domestic clades, which can be interpreted as individual introductions, confirmed that A.23.1 with E484K (1 clade) probably has a domestic origin as no genomes of the same clade were observed internationally ([Methods](#), Fig. 5d and Extended Data Fig. 7). The estimated number of introductions was lowest for B.1.1.318 (3 introductions, range = 1–6), and highest for

Beta (49 introductions, range = 45–58) and Eta (30 introductions, range = 18–34). Although our data exclude genomes sampled directly from travellers, these repeated introductions show that the true rate of transmission is lower than the observed increase in the number of surveillance genomes.

## The rise of Delta from April to June 2021

The B.1.617.1/Kappa and B.1.617.2/Delta lineages, which were first detected in India in 2020, first appeared in English surveillance samples in March 2021. In contrast to other VOCs, Delta/Kappa do not contain N501Y or E484K mutations, but their L452R mutation may reduce antibody recognition<sup>27</sup> and P681R enhances furin cleavage<sup>30</sup>, similar to the P681H mutation of Alpha. The frequency of Delta, which harbours further spike mutations of unknown function, increased rapidly and reached levels of 98% (12,474/12,689) on 26 June 2021 (Fig. 5a, b). Although initially constrained to a small number of large local clusters, such as in Bolton, in May 2021 (Fig. 5c), Delta was detected in all LTLAs by 26 June 2021 (Fig. 1c). The sweep of Delta occurred at a rate of around 59% (growth per 5.1 d, CI = 53–66) higher than Alpha with minor regional variation (Fig. 2a, Extended Data Fig. 4e and Supplementary Table 4).

The rapid rise of Delta contrasts with Kappa, which grew more slowly despite being introduced at a similar time and into a similar demographic background (Figs. 2a and 5b). This is also evident in the phylogeographical analysis (based on data as of 1 May 2021). The 224 genomes of Delta derive from larger clades (23 introductions, range = 6–40; around 10 genomes for every introduction) compared with the 80 genomes of Kappa (17 introductions, range = 15–31; around 3–4 genomes per introduction) and also other variants (Fig. 5d and Extended Data Fig. 8). The AY.1 lineage, derived from Delta and containing an additional K417N mutation, appeared only transiently (Fig. 5b).

The sustained domestic growth of Delta and its international spread<sup>31</sup> relative to the Alpha lineage are the first evidence of a biological growth advantage. The causes appear to be a combination of increased transmissibility and immune evasion. Evidence for higher transmissibility includes the fast growth in younger unvaccinated age groups, reports of elevated secondary attack rates<sup>32</sup> and a higher viral load<sup>33</sup>. Furthermore, vaccine efficacy against infection by Delta is diminished, depending on the type of vaccine<sup>34,35</sup>, and reinfection is more frequent<sup>36</sup>, both supported by experimental research demonstrating the reduced antibody neutralization of Delta by vaccine-derived and convalescent sera<sup>37,38</sup>.

The higher growth rate of Delta—combined with gradual reopening and proceeding vaccination—repeated the dichotomous pattern of lineage-specific decline and growth, although now with declining Alpha ( $R_t < 1$ ) and growing Delta ( $R_t > 1$ ; Fig. 5e, f).

Overall, we estimate that the spread of more transmissible variants between August 2020 and early summer 2021 increased the average growth rate of circulating SARS-CoV-2 in England by a factor of 2.39 (95% CI = 2.25–2.42; Fig. 5g). Thus, previously effective interventions may prove to be insufficient to contain newly emerging and more transmissible variants.

## Discussion

Our dense genomic surveillance analysis identified lineages that consistently grew faster than others in each local authority and, therefore, at the same time, under the same restrictions and in a comparable population. This pinpointed a series of variants with elevated transmissibility, in broad agreement with other reports<sup>10,11,13,15,31</sup>.

However, a number of limitations exist. The growth rates of rare new variants are stochastic due to introductions and superspreading. Local outbreaks of the Beta and Gamma variants triggered asymptomatic surge testing, which may have reduced their spread. Furthermore, transmission depends both on the viral variant and the immunity of the host population, which changed from less than 20% to over 90% in the study period<sup>39</sup>. This will influence the growth rates of variants with immune evasion capabilities over time. The effect of immunity is currently not modelled, but may become more important in the future as SARS-CoV-2 becomes endemic. Further limitations are discussed in the Limitations section of the [Methods](#).

The third and fourth waves in England were each caused by more transmissible variants, which outgrew restrictions that were sufficient to suppress previous variants. During the second national lockdown, Alpha grew despite falling numbers for other lineages and, similarly, Delta took hold in April and May when cases of Alpha were declining. The fact that such growth was initially masked by the falling cases of dominant lineages highlights the need for dense genomic surveillance and rapid analysis to devise optimal and timely control strategies. Such surveillance should ideally be global as, even though Delta was associated with a large wave of cases in India, its transmissibility remained unclear at the time due to a lack of systematic genomic surveillance data.

The 2.4-fold increase in growth rate during the study period as a result of new variants is also likely to have consequences for the future course of the pandemic. If this increase in growth rate was explained solely by higher transmissibility, it would raise the basic reproduction number  $R_0$  from a value of around 2.5–3 in spring 2020 (ref. 40) to the range of 6–7 for Delta. This is likely to spur new waves of the epidemic in countries that have to date been able to control the epidemic despite low vaccination rates, and it may exacerbate the situation elsewhere. Although the exact herd-immunity threshold depends on contact patterns and the distribution of immunity

across age groups<sup>41,42</sup>, it is worth considering that Delta may increase the threshold to values around 0.85. Given current estimates of vaccine efficacy<sup>34,35,43</sup> this would require nearly 100% vaccination coverage. Even though more than 90% of adults had antibodies against SARS-CoV-2 (ref. <sup>39</sup>) and close to 70% had received two doses of vaccination, England saw rising Delta variant cases in the first weeks of July 2021. It can therefore be expected that other countries with high vaccination coverage are also likely to experience rising cases when restrictions are lifted.

SARS-CoV-2 is likely to continue its evolutionary adaptation process to humans<sup>44</sup>. To date, variants with considerably higher transmissibility have had strongest positive selection, and swept through England during the 10 months of this investigation. However, the possibility that an increasingly immune population may now select for variants with better immune escape highlights the need for continued systematic and, ideally, global genomic surveillance.

## Methods

### Pillar 2 SARS-CoV-2 testing data

Publicly available daily SARS-CoV-2 test result data from testing for the wider population outside the National Health Service (Pillar 2 newCasesBySpecimenDate) were downloaded from <https://coronavirus.data.gov.uk/> spanning the date range from 1 September 2020 to 30 June 2021 for 315 English LTLAs (downloaded on 20 July 2021). These data are mostly positive PCR tests, with about 4% of results from lateral flow tests without PCR confirmation. In this dataset, the City of London is merged with Hackney, and the Isles of Scilly are merged with Cornwall due to their small number of inhabitants, thereby reducing the number of English LTLAs from 317 to 315. Population data for each LTLA were downloaded from the Office of National Statistics (ONS; <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforkenglandandwalesscotlandandnorthernireland>).

### SARS-CoV-2 surveillance sequencing

In total, 281,178 tests (September 2020 to June 2021) were collected as part of random surveillance of positive tests of residents of England from four Pillar 2 Lighthouse laboratories. The samples were collected between 1 September 2020 and 26 June 2021. A random selection of samples was taken, after excluding those that were known to be taken during quarantine of recent travellers, and samples from targeted and local surge testing efforts. The available metadata made this selection imperfect,

but these samples should be an approximately random selection of infections in England during this time period, and the large sample size makes our subsequent inferences robust.

We amplified RNA extracts from these tests with  $C_t < 30$  using the ARTIC amplicon protocol (<https://www.protocols.io/workspaces/coguk/publications>). We sequenced 384-sample pools on Illumina NovaSeq, and produced consensus fasta sequences according to the ARTIC nextflow processing pipeline (<https://github.com/connor-lab/ncov2019-artic-nf>). Lineage assignments were made using Pangolin<sup>5</sup>, according to the latest lineage definitions at the time, except for B.1.617, which we reanalysed after the designation of sublineages B.1.617.1, B.1.617.2 and B.1.617.3. Lineage prevalence was computed from 281,178 genome sequences. The genomes were mapped to the same 315 English LTLAs as for the testing data described above. Mapping was performed from outer postcodes to LTLA, which can introduce some misassignment to neighbouring LTLAs. Furthermore, lineages in each LTLA were aggregated to counts per week for a total of 43 weeks, defined beginning on Sunday and ending on Saturday.

Finally, the complete set of 328 SARS-CoV-2 PANGO lineages was collapsed into  $l = 71$  lineages using the underlying phylogenetic tree, such that each resulting lineage constituted at least 100 genomes, unless the lineage has been designated a VOC, variant under investigation (VUI) or variant in monitoring by Public Health England<sup>32</sup>.

## Spatiotemporal genomic surveillance model

A hierarchical Bayesian model was used to fit local incidence data in a given day in each local authority and jointly estimate the relative historical prevalence and transmission parameters. In the following,  $t$  denotes time and is measured in days. We use the convention that bold lowercase symbols, such as  $\mathbf{b}$ , indicate vectors.

## Motivation

Suppose that  $\{\{\bf{x}\}\}^{\prime}(t) = (\{\bf{b}\} + \{\bf{r}_0(t)\}) \cdot \{\bf{x}\}(t)$  describes the ordinary differential equation (ODE) for the viral dynamics for a set of  $l$  different lineages. Here  $r_0(t)$  is a scalar time-dependent logarithmic growth rate that is thought to reflect lineage-independent transmission determinants, which changes over time in response to behaviour, non-pharmaceutical interventions (NPIs) and immunity. This reflects a scenario in which the lineages differ only in terms of the intensity of transmission, but not the intergeneration time distribution. The ODE is solved by  $\{\bf{x}\}(t) = \{\bf{e}\}^{\wedge}(\{\bf{c}\} + \{\bf{b}\}t + \int_{0}^{t} \{\bf{r}_0(u)\} du)$ . The term  $v(t)$  contributes the same factor to each lineage and therefore drops from the relative proportions of

lineages  $\frac{\sum \mathbf{x}(t)}{\sum \mathbf{x}(t) + \mathbf{b}t}$ .

In the given model, the lineage prevalence  $\mathbf{p}(t)$  follows a multinomial logistic-linear trajectory. Moreover, the total incidence factorizes into  $\nu(t) = \sum \mathbf{e}^T \mathbf{c} + \mathbf{b}t$ , which provides a basis to separately estimate the total incidence  $\mu(t)$  from Pillar 2 test data and lineage-specific prevalence  $\mathbf{p}(t)$  from genomic surveillance data (which are taken from a varying proportion of positive tests). By using the equations above, one can subsequently calculate lineage-specific estimates by multiplying  $\mu(t)$  with the respective genomic proportions  $\mathbf{p}(t)$ .

## Incidence

In the following text, we describe a flexible semi-parametric model of the incidence. Let  $\mu(t)$  be the expected daily number of positive Pillar 2 tests and  $s$  the population size in each of 315 LTLAs. Denote  $\lambda(t) = \log(\mu(t)/s)$  the logarithmic daily incidence per capita at time  $t$  in each of the 315 LTLAs.

Suppose  $f(t)$  is the daily number of new infections caused by the number of people infected at time  $t$ . As new cases are noticed and tested only after a delay  $u$  with distribution  $g$ , the observed number of cases  $f^*(t)$  will be given by the convolution

$$f^*(t) = \int_0^\infty g(u)f(t-u)du = (g * f)(t).$$

The time from infection to test is given by the incubation time plus the largely unknown distribution of the time from symptoms to test, which, in England, was required to take place within 5 d of symptom onset. To account for these factors, the log normal incubation time distribution from ref. [46](#) is scaled by the equivalent of changing the mean by 2 d. The convolution shifts cases approximately 6 d into the future and also spreads them out according to the width of  $g$  (Extended Data Fig. [2a](#)).

To parametrize the short- and longer-term changes of the logarithmic incidence  $\lambda(t)$ , we use a combination of  $h$  weekly and  $k-h$  monthly cubic basis splines ( $\mathbf{f}(t) = (\mathbf{f}_1(t), \dots, \mathbf{f}_{k-h}(t))$ ). The knots of the  $h$  weekly splines uniformly tile the observation period except for the last 6 weeks.

Each spline basis function is convolved with the time to test distribution  $g$ ,  $(\mathbf{f} * g)(t) = (\mathbf{f}_1 * g)(t), \dots, (\mathbf{f}_{k-h} * g)(t)$  as outlined above and used to fit the logarithmic incidence. The derivatives of the original basis  $\mathbf{f}(t)$  are used to calculate the underlying growth rates and  $R_t$  values, as shown further below. The convolved spline basis  $\mathbf{f}^*(t)$  is used to fit the per capita incidence in each LTLA as (Extended Data Fig. [2b](#)):

$$\$ \$ \boldsymbol{\lambda}(t) = \mathbf{B} \times \boldsymbol{f}^{\ast}(t). \$ \$$$

This implies that fitting the incidence function for each of the  $m$  local authorities is achieved by a suitable choice of coefficients  $\mathbf{B}$  in  $\mathbf{R}^m$ , that is one coefficient for each spline function for each of the LTAs. The parameters  $\mathbf{B}$  have a univariate normal prior distribution each, which reads for LTLA  $i$  and spline  $j$ :

$$\$ \$ \mathbf{B}_{ij} \sim N(0, \sigma_j). \$ \$$$

The s.d. of the prior regularizes the amplitude of the splines and is chosen as  $\sigma_j=0.2$  for weekly splines and  $\sigma_j=1$  for monthly splines. This choice was found to reduce the overall variance resulting from the high number of weekly splines, meant to capture rapid changes in growth rates, but which can lead to instabilities particularly at the end of the time series, when not all effects of changes in growth rates are observed yet. The less regularized monthly splines reflect trends on the scale of several weeks and are therefore subject to less noise.

Finally, we introduce a term accounting for periodic differences in weekly testing patterns (there are typically 30% lower specimens taken on weekends; Fig. 1a):

$$\$ \$ \mathbf{\mu} \mid \mathbf{B} \sim \mathbf{\mu}(t) \cdot \delta(t), \$ \$$$

where the scalar  $\delta(t) = \delta(t-i \times 7)$ ,  $i \in \mathbb{N}$  and prior distribution  $\delta(t) \sim \text{rm{L}}(\text{rm{o}}) \text{rm{g}}(\text{rm{N}}) \text{rm{o}}(\text{rm{r}}) \text{rm{m}}(\text{rm{a}}) \text{rm{l}}(0,1)$  for  $t=1, \dots, 6$  and  $\delta(0)=1$ .

The total incidence was fitted to the observed number of positive daily tests  $\mathbf{X}$  by a negative binomial with a dispersion  $\omega=10$ . The overdispersion buffers against non-Poissonian uncorrelated fluctuations in the number of daily tests.

$$\$ \$ \mathbf{X}(t) \sim \text{rm{N}} \mathbf{B} \mathbf{\mu}(t) \text{limits}^{\sim}(t), \omega. \$ \$$$

The equation above assumes that all elements of  $\mathbf{X}(t)$  are independent, conditional on  $\tilde{\boldsymbol{\mu}}(t)$ .

## Growth rates and $R_t$ values

A convenient consequence of the spline basis of  $\log(\boldsymbol{\mu}) = \boldsymbol{\lambda}$ , is that the delay-adjusted daily logarithmic growth rate  $\mathbf{r}(t) = \boldsymbol{\lambda}'(t)$  of the local epidemic simplifies to:

$$\$ \$ \{ \bf{r} \}(t) = \{ \bf{B} \} \times \{ \bf{f} \} \}^{\prime}(t), \$ \$$$

where  $\{ \bf{f} \}_j^{\prime}(t)$  represents the first derivative of the  $j$ th cubic spline basis function.

To express the daily growth rate as an approximate reproductive number  $R_t$ , one needs to consider the distribution of the intergeneration time, which is assumed to be gamma distributed with mean 6.3 d ( $\alpha = 2.29$ ,  $\beta = 0.36$ )<sup>46</sup>. The  $R_t$  value can be expressed as a Laplace transform of the intergeneration time distribution<sup>47</sup>. Effectively, this shortens the relative time period because the exponential dynamics put disproportionately more weight on stochastically early transmissions over late ones. For reasons of simplicity and being mindful also of the uncertainties of the intergeneration time distribution, we approximate  $R_t$  values by multiplying the logarithmic growth rates with a value of  $\{ \bar{\tau} \}_{\rm e} = 5.1$  d, which was found to be a reasonable approximation to the convolution required to calculate  $R_t$  values (denoted here by the lower case symbol  $\{ \boldsymbol{\rho} \}(t)$ ) in line with our convention for vector-variate symbols and to avoid confusion with the epidemiological growth rate  $r_t$ ,

$$\$ \$ \log(\{ \boldsymbol{\rho} \}(t)) \approx \frac{d}{dt} \log(\{ \boldsymbol{\mu} \}(t)) = \{ \bf{r} \}(t) \{ \bar{\tau} \}_{\rm e} \$ \$$$

Thus, the overall growth rate scaled to an effective inter generation time of 5.1 d can be readily derived from the derivatives of the spline basis and the corresponding coefficients. The values derived from the approach are in very close agreement with those of the method of ref. <sup>48</sup>, but shifted according to the typical delay from infection to test (Extended Data Fig. 2b).

## Genomic prevalence

The dynamics of the relative frequency  $\mathbf{P}(t)$  of each lineage was modelled using a logistic-linear model in each LTLA, as described above. The logistic prevalence of each lineage in each LTLA is defined as  $\{ \bf{L} \}(t) = \{ \rm{l} \} \{ \rm{o} \} \{ \rm{g} \} \{ \rm{i} \} \{ \rm{t} \} (\{ \bf{P} \}(t))$ . This is modelled using the piecewise linear expression

$$\$ \$ \{ \bf{L} \}(t) = \{ \bf{C} \} + \{ \bf{b} \} \cdot \{ \bf{t} \}_{+}, \$ \$$$

where  $\mathbf{b}$  may be interpreted as a lineage-specific growth advantage and  $\mathbf{C}$  as an offset term of dimension (LTLA  $\times$  lineages). Time  $\{ \bf{t} \}_{+}$  is measured since introduction  $t_0$  and is defined as

$\$ \$ \{ \{ \bf{t} \} \} _{+} = t - \{ \{ \bf{t} \} \} _{0} \backslash, \{ \rm{i} \} \{ \rm{f} \} \backslash, t > \{ \{ \bf{t} \} \} _{0} \backslash, \{ \rm{e} \} \}$   
 $\{ \rm{l} \} \{ \rm{s} \} \{ \rm{e} \} - \{ \rm{infty} \} \$ \$$

and accounts for the fact that lineages can be entirely absent prior to a stochastically distributed time period preceding their first observation. This is because, in the absence of such a term, the absence of a lineage prior to the point of observation can only be explained by a higher growth rate compared with the preceding lineages, which may not necessarily be the case. As the exact time of introduction is generally unknown, a stochastic three-week period of  $\sim \{ \rm{Unif} \} (-14,0) + \{ \{ \bf{t} \} \} _{0} ^{\{ \rm{obs} \}}$  prior to the first observation  $\{ \{ \bf{t} \} \} _{0} ^{\{ \rm{obs} \}}$  was chosen.

As the inverse logit transformation projects onto the  $l-1$  dimensional simplex  $\{ \{ \rm{S} \} _{l-1} \}$  and therefore loses one degree of freedom, B.1.177 was set as a baseline with

$\$ \$ \{ \{ \bf{L} \} \} _{\cdot} \{ \bf{0} \}(t) = 0. \$ \$$

The offset parameters  $C$  are modelled across LTAs as independently distributed multivariate normal random variables with a lineage-specific mean  $\mathbf{c}$  and covariance  $\Sigma = 10 \cdot \{ \rm{I} \} _{l-1}$ , where  $\{ \rm{I} \} _{l-1}$  denotes an  $(l-1) \times (l-1)$  identity matrix. The lineage-specific parameters growth rate  $\mathbf{b}$  and average offset  $\mathbf{c}$  are modelled using IID Normal prior distributions

$\$ \$ \{ \bf{b} \} \sim N(0,0.2) \$ \$$   
 $\$ \$ \{ \bf{c} \} \sim N(-10,5) \$ \$$

The time-dependent relative prevalence  $\mathbf{P}(t)$  of SARS-CoV2 lineages was fitted to the number of weekly genomes  $\mathbf{Y}(t)$  in each LTA by a Dirichlet-multinomial distribution with expectation  $\mathbb{E}[\mathbf{Y}(t)] \approx \mathbf{P}(t) \cdot \mathbf{G}(t)$  where  $\mathbf{G}(t)$  are the total number of genomes sequenced from each LTA in each week. For LTA  $i$ , this is defined as:

$\$ \$ \{ \{ \bf{Y} \} \} _i \cdot (t) \sim \{ \rm{D} \} \{ \rm{i} \} \{ \rm{r} \} \{ \rm{M} \} \{ \rm{u} \}$   
 $\{ \rm{l} \} \{ \rm{t} \} (\alpha _0 + \{ \boldsymbol{\alpha} \} _1) \{ \bf{P} \} _i \cdot (t), \{ \bf{G} \} _i(t). \$ \$$

The scalar parameter  $\alpha _0 = 0.01$  can be interpreted as a weak prior with expectation  $1/n$ , making the model less sensitive to the introduction of single new lineages, which can otherwise exert a very strong effect. Furthermore, the array  $\{ \boldsymbol{\alpha} \} _1 = \frac{1}{2} \{ \rm{cases} \}$  increases the variance to account for the fact that, especially at high sequencing coverage (genomes  $\approx$  cases), cases and therefore genomes are likely to be correlated and overdispersed as

they may derive from a single transmission event. Other choices such as  $\{\{\boldsymbol{\alpha}\}\}_{1}=1,000$ , which make the model converge to a standard multinomial, leave the conclusions qualitatively unchanged. This model aspect is illustrated in Extended Data Fig. 2c.

## Lineage-specific incidence and growth rates

From the two definitions above it follows that the lineage-specific incidence is given by multiplying the total incidence in each LTLA  $\mu(t)$  with the corresponding lineage frequency estimate  $\mathbf{P}(t)$  for lineage  $j$  at each time point

$$\{\{\{\mathbf{M}\}\}\}_{\cdot j}(t) = \{\boldsymbol{\mu}\}(t) \cdot \{\{\mathbf{P}\}\}_{\cdot j}(t) \quad \text{for } (j=0, \dots, l-1)$$

Further corresponding lineage-specific  $R_t$  values  $\mathbf{R}(t)$  in each LTLA can be calculated from the lineage-agnostic average  $R_t$  value  $\rho(t)$  and the lineage proportions  $\mathbf{P}(t)$  as

$$\log \{\mathbf{R}\}(t) = \log \rho(t) + \bar{\tau} \{e\} - \mathbf{P}(t) \cdot \mathbf{b}$$

By adding the log-transformed growth rate fold changes  $\mathbf{b}$  and subtracting the average log-transformed growth rate change  $(\{\mathbf{P}(t)\} \cdot \mathbf{b})$ , it follows that  $\{\{\mathbf{R}\}\}_{i,j}(t) = \{\{\mathbf{R}\}\}_{i,0}(t) \cdot \bar{\tau} \{e\}^{\{\mathbf{b}\}} / \{\mathbf{b}\}$ , where  $\{\{\mathbf{R}\}\}_{i,0}(t)$  is the  $R_t$  value of the reference lineage  $j=0$  (for which  $\{\{\mathbf{b}\}\}_{0}=0$ ) in LTLA  $i$ . It follows that all other lineage-specific the  $R_t$  values are proportional to this baseline at any given point in time with factor  $\{\bar{\tau} \{e\}\}^{\{\mathbf{b}\}}$ .

## Inference

The model was implemented in numpyro<sup>49,50</sup> and fitted using stochastic variational inference<sup>51</sup>. Guide functions were multivariate normal distributions for each row (corresponding to an LTLA) of  $\mathbf{B}$ ,  $\mathbf{C}$  to preserve the correlations across lineages and time as well as for  $(\mathbf{b}, \mathbf{c})$  to also model correlations between growth rates and typical introduction.

## Phylogeographic analyses

To infer VOC introduction events into the UK and corresponding clade sizes, we investigated VOC genome sequences from GISAID (<https://www.gisaid.org/>) available from any country. We downloaded multiple sequence alignments of genome sequences with the release dates 17 April 2021 (for the analysis of the lineages A.23.1,

B.1.1.318, B.1.351 and B.1.525) and 5 May 2021 (for the analysis of the B.1.617 sublineages). We next extracted a subalignment from each lineage (according to the 1 April 2021 version of PANGOLin for the 17 April 2021 alignment and the 23 April 2021 version of PANGOLin for the 5 May 2021 alignment) and, for each subalignment, we inferred a phylogeny through maximum likelihood using FastTree2 (v.2.1.11)<sup>52</sup> with the default options and GTR substitution model<sup>53</sup>.

On each VOC/VUI phylogeny, we inferred the minimum and maximum number of introductions of the considered SARS-CoV-2 lineage into the UK compatible with a parsimonious migration history of the ancestors of the considered samples; we also measured clade sizes for one specific example parsimonious migration history. We counted only introduction events into the UK that resulted in at least one descendant from the set of UK samples that we considered in this work for our hierarchical Bayesian model; similarly, we measured clade sizes by the number of UK samples considered here included in such clades. Multiple occurrences of identical sequences were counted as separate cases, as this helped us to identify rapid SARS-CoV-2 spread.

When using parsimony, we considered only migration histories along a phylogenetic tree that are parsimonious in terms of the number of migration events from and to the UK (in practice, we collapse all of the non-UK locations into a single one).

Furthermore, as SARS-CoV-2 phylogenies present substantial numbers of polytomies, that is, phylogenetic nodes where the tree topology cannot be reconstructed due to a lack of mutation events on certain branches, we developed a tailored dynamic programming approach to efficiently integrate over all possible splits of polytomies and over all possible parsimonious migration histories. The idea of this method is somewhat similar to typical Bayesian phylogeographic inference<sup>54</sup> in that it enables us to at least in part integrate over phylogenetic uncertainty and uncertainty in migration history; however, it also represents a very simplified version of these analyses, more so than ref. <sup>16</sup>, as it considers most of the phylogenetic tree as fixed, ignores sampling times and uses parsimony instead of a likelihood-based approach. Parsimony is expected to represent a good approximation in the context of SARS-CoV-2, due to the shortness (both in time and substitutions) of the phylogenetic branches considered<sup>55,56</sup>. The main advantage of our approach is that, owing to the dynamic programming implementation, it is more computationally efficient than Bayesian alternatives, as the most computationally demanding step is the inference of the maximum likelihood phylogenetic tree. This enables us to infer plausible ranges for numbers of introduction events for large datasets and to quickly update our analyses as new sequences become available. The other advantage of this approach is that it enables us to easily customize the analysis and to focus on inferred UK introductions that result in at least one UK surveillance sample, while still making use of non-surveillance UK samples to inform the inferred phylogenetic tree and migration history. Note that possible biases due to

uneven sequencing rates across the world<sup>55</sup> apply to our approach as well as other popular phylogeographic methods. Our approach works by traversing the maximum likelihood tree starting from the terminal nodes and ending at the root (postorder traversal). Here, we define a ‘UK clade’ as a maximal subtree of the total phylogeny for which all terminal nodes are from the UK, all internal nodes are inferred to be from the UK and at least one terminal node is a UK surveillance sample; the size of a UK clade is defined as the number of UK surveillance samples in it. At each node, using values already calculated for all children nodes (possibly more than two children in the case of a multifurcation), we calculate the following quantities: (1) the maximum and minimum number of possible descendant UK clades of the current node, over the space of possible parsimonious migration histories, and conditional on the current node being UK or non-UK; (2) the number of migration events compatible with a parsimonious migration history in the subtree below the current node, and conditional on the current node being UK or non-UK; (3) the size so far of the UK clade the current node is part of, conditional on it being UK; and (4) a sample of UK clade sizes for the subtree below the node. To calculate these quantities, for each internal node, and conditional on each possible node state (UK or non-UK), we consider the possible scenarios of having 0 or 1 migration events between the internal node and its children nodes (migration histories with more than 1 migration event between the node and its children are surely not parsimonious in our analysis and can be ignored).

To confirm the results of our analyses based on parsimony, we also used the new Bayesian phylogenetic approach Thorney BEAST<sup>16</sup> ([https://beast.community/thorney\\_beast](https://beast.community/thorney_beast)) for VOCs for which it was computationally feasible, that is, excluding B.1.351. For each VOC, we used in Thorney BEAST the same topology inferred with FastTree2 as for our parsimony analysis; we also used treetime<sup>57</sup> v.0.8.2 to estimate a timed tree and branch divergences for use in Thorney BEAST. We used a two-state (UK and non-UK) migration model<sup>54</sup> of migration to infer introductions into the UK but again counted, from the posterior sample trees, only UK clades with at least one UK surveillance sample. We used a Skygrid<sup>58</sup> tree coalescent prior with six time intervals. The comparison of parsimony and Bayesian estimates is shown in Extended Data Fig. 8d.

## ONS infection survey analysis

Data from the cross-sectional infection survey were downloaded from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionssurveypilot/30april2021>.

Comparison of ONS incidence estimates with hospitalization, case and death rates was conducted by estimating infection trajectories separately from observed cases, hospitalizations and deaths<sup>59,60</sup>, convolving them with estimated PCR detection

curves<sup>61</sup>, and dividing the resulting PCR prevalence estimates by the estimated prevalence from the ONS Community Infection Survey at the midpoints of the two-week intervals over which prevalence was reported in the survey.

## Maps

Maps were plotted using LTLA shapefiles (<https://geoportal.statistics.gov.uk/datasets/69dc11c7386943b4ad8893c45648b1e1>), sourced from the ONS, which is licensed under the Open Government Licence v.3.0.

## Limitations

A main limitation of the analysis is that the transmission model is deterministic, whereas the spread of variants is a stochastic process. Although the logistic growth assumption is a consistent estimator of the average transmission dynamics, individual outbreaks may deviate from these averages and therefore produce unreliable estimates.

Stochastic growth effects are accounted for only in terms of (uncorrelated) overdispersion and the offset at the time of the introduction. For these reasons, the estimated growth rates may not accurately reflect the viral transmissibility, especially at a low prevalence. It is therefore important to assess whether consistent growth patterns in multiple independent areas are observed. We note that the posterior distribution of the growth rates of rare variants tends to be biased to the baseline due to the centred prior.

In its current form, the model accounts for only a single introduction event per LTLA. Although this problem is in part alleviated by the high spatial resolution, which spreads introductions across 315 LTLAs, it is important to investigate whether sustained introductions inflate the observed growth rates, as in the case of the Delta variant or other VOCs and VUIs. This can be achieved by a more detailed phylogeographic assessment and through the assessment of monophyletic sublineages.

Furthermore, there is no explicit transmission modelled from one LTLA to another. As each introduction is therefore modelled separately, this makes the model conservative in ascertaining elevated transmission as single observed cases across different LTLAs can be explained by their introduction.

The inferred growth rates also cannot identify a particular mechanism of altered transmission. Biological mechanisms include a higher viral load, longer infectivity or greater susceptibility. Lineages could potentially differ by their intergeneration time, which would lead to nonlinear scaling. Here we did not find convincing evidence in incidence data for such effects, in contrast to previous reports<sup>23</sup>. However, contact-

tracing data indicate that the intergeneration time may be shortening for more transmissible lineages such as Delta<sup>33,62</sup>. Cases of the Beta and Gamma VOCs may have been more intensely contact traced and triggered asymptomatic surge testing in some postcode areas. This may have reduced the observed growth rates relative to other lineages.

Lineages, such as Beta, Gamma or Delta also differ in their ability to evade previous immunity. As immunity changes over time, this might lead to a differential growth advantage over time. It is therefore advisable to assess whether a growth advantage is constant over periods in which immunity changes considerably.

A further limitation underlies the nature of lineage definition and assignment. The PANGO lineage definition<sup>5</sup> assigns lineages to geographical clusters, which have by definition expanded, and this can induce a certain survivor bias, often followed by winner's curse. Another issue results from the fact that very recent variants may not be classified as a lineage despite having grown, which can inflate the growth rate of ancestral lineages over sublineages.

As the total incidence is modelled on the basis of the total number of positive PCR tests, it may be influenced by testing capacity; the total number of tests approximately tripled between September 2020 and March 2021. This can potentially lead to a time trend in recorded cases and therefore baseline  $R_t$  values if the access to testing changed, for example, by too few tests being available during periods of high incidence, or changes to the eligibility to intermittently test with fewer symptoms. Generally, the observed incidence was in good agreement with representative cross-sectional estimates from the ONS<sup>63,64</sup>, except for a period of peak incidence from late December 2020 to January 2021 (Extended Data Fig. 1d). Values after 8 March 2021 need to be interpreted with caution as Pillar 2 PCR testing was supplemented by lateral flow devices, which increased the number of daily tests to more than 1.5 million. Positive cases were usually confirmed by PCR and counted only once.

The modelled curves are smoothed over intervals of approximately 7 d using cubic splines, creating the possibility that later time points influence the period of investigation and cause a certain waviness of the  $R_t$  value pattern. An alternative parameterization using piecewise linear basis functions per week (that is, constant  $R_t$  values per week) leaves the overall conclusions and extracted parameters broadly unchanged.

## Ethical approval

This study was performed as part of surveillance for COVID-19 under the auspices of Section 251 of the National Health Service Act 2006. It therefore did not require

individual patient consent or ethical approval. The COG-UK study protocol was approved by the Public Health England Research Ethics Governance Group.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

PCR test data are publicly available online (<https://coronavirus.data.gov.uk/>). A filtered, privacy conserving version of the lineage–LTLA–week dataset is publicly available online (<https://covid19.sanger.ac.uk/downloads>) and enables strong reproduction of our results, despite a small number of cells having been suppressed to avoid disclosure. Full SARS-CoV-2 genome data and geolocations can be obtained under controlled access from <https://www.cogconsortium.uk/data/>. Application for full data access requires a description of the planned analysis and can be initiated at coguk\_DataAccess@medschl.cam.ac.uk. The data and a version of the analysis with fewer lineages can be interactively explored at <https://covid19.sanger.ac.uk>. [Source data](#) are provided with this paper.

## Code availability

The genomic surveillance model is implemented in Python and available at GitHub (<https://github.com/gerstung-lab/genomicsurveillance>) and as a PyPI package (genomicsurveillance). Specific code for the analyses of this study can be found as individual Google colab notebooks in the same repository. These were run using Python v.3.7.1 (packages: matplotlib (v.3.4.1), numpy (v.1.20.2), pandas (v.1.2.3), scikit-learn (v.0.19.1), scipy (v.1.6.2), seaborn (v.0.11.1), jax (v.0.2.8), genomicsurveillance (v.0.4.0), numpyro (v.0.4.0)). The phylogeographic analyses were performed using Thorney Beast (v.0.1.1) and <https://github.com/NicolaADM/phylogeographySARS-CoV-2>. Code for the ONS infection survey analysis is available at GitHub ([https://github.com/jhellewell14/ons\\_severity\\_estimates](https://github.com/jhellewell14/ons_severity_estimates)).

## References

1. 1.

Rambaut, A. *Phylogenetic Analysis of nCoV-2019 Genomes* (Virological, 2020); <https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356>

2. 2.

Nextstrain Team *Genomic Epidemiology of Novel Coronavirus—Global Subsampling* (Nextstrain, 2020); <https://nextstrain.org/ncov/global?l=clock>

3. 3.

Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

4. 4.

Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75 (2021).

5. 5.

Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

6. 6.

O'Toole, Á. et al. *Global Report Investigating Novel Coronavirus Haplotypes* [https://cov-lineages.org/global\\_report.html](https://cov-lineages.org/global_report.html) (2021).

7. 7.

Hodcroft, E. B. et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **595**, 707–712 (2021).

8. 8.

Tegally, H., Wilkinson, E., Lessells, R.J. et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* **27**, 440–446 (2021).

9. 9.

Rambaut, A. et al. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations* (Virological, 2020); <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>

10. 10.

Volz, E., Mishra, S., Chand, M. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).

11. 11.

Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).

12. 12.

O'Toole, Á. et al. *Tracking the International Spread of SARS-CoV-2 Lineages B.1.1.7 and B.1.351/501Y-V2* (Virological, 2021);  
<https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592>

13. 13.

Washington, N. L. et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**, 2587–2594 (2021).

14. 14.

Faria, N. R. et al. *Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in Manaus: Preliminary Findings* (Virological, 2021);  
<https://www.icpcovid.com/sites/default/files/2021-01/Ep%20102-1%20Genomic%20characterisation%20of%20an%20emergent%20SARS-CoV-2%20lineage%20in%20Manaus%20Genomic%20Epidemiology%20-%20Virological.pdf>

15. 15.

Faria, N. R. et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **372**, 815–21 (2021).

16. 16.

du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).

17. 17.

Danish Covid-19 Genome Consortium *Genomic Overview of SARS-CoV-2 in Denmark* (2021); <https://www.covid19genomics.dk/statistics>

18. 18.

Kraemer, M. U. G. et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science* **373**, 889–895 (2021).

19. 19.

Park, S. W. et al Roles of generation-interval distributions in shaping relative epidemic strength, speed, and control of new SARS-CoV-2 variants. Preprint at *medRxiv* <https://doi.org/10.1101/2021.05.03.21256545> (2021).

20. 20.

Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 (2020).

21. 21.

Zahradník, J., Marciano, S., Shemesh, M. et al. SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nat. Microbiol.* **6**, 1188–1198 (2021).

22. 22.

Brown, J. C. et al. Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 202012/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.24.432576> (2021).

23. 23.

Vöhringer, H. et al. *Lineage-specific Growth of SARS-CoV-2 B.1.1.7 During the English National Lockdown* (Virological, 2020); <https://virological.org/t/lineage-specific-growth-of-sars-cov-2-b-1-1-7-during-the-english-national-lockdown/575/2>

24. 24.

The Health Protection (Coronavirus, Restrictions) (All Tiers) (England) Regulations 2020. *Wikipedia* [https://en.wikipedia.org/w/index.php?title=The\\_Health\\_Protection\\_\(Coronavirus,\\_Restrictions\)\\_\\_\(All\\_Tiers\)\\_\\_\(England\)\\_Regulations\\_2020&oldid=1014831173](https://en.wikipedia.org/w/index.php?title=The_Health_Protection_(Coronavirus,_Restrictions)__(All_Tiers)__(England)_Regulations_2020&oldid=1014831173) (2021).

25. 25.

Steel, K. & Davies, B. *Coronavirus (COVID-19) Infection Survey, Antibody and Vaccination Data for the UK* (ONS, 2021);  
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveyantibodydatafortheuk/28april2021>

26. 26.

Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe* **29**, 463-476 (2021).

27. 27.

Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).

28. 28.

Zhou, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361 (2021).

29. 29.

Planas, D. et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat. Med.* **27**, 917–924 (2021).

30. 30.

Peacock, T. P. et al. The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.28.446163> (2021).

31. 31.

Campbell, F. et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Euro Surveill.* **26**, 2100509 (2021).

32. 32.

*Investigation of Novel SARS-CoV-2 Variants of Concern* Technical briefing 10  
(Public Health England, 2021);  
<https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201>

33. 33.

Li, B. et al. Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant. Preprint at *medRxiv*  
<https://doi.org/10.1101/2021.07.07.21260122> (2021).

34. 34.

Nasreen, S. et al. Effectiveness of COVID-19 vaccines against variants of concern in Ontario, Canada. Preprint at *medRxiv*  
<https://doi.org/10.1101/2021.06.28.21259420> (2021).

35. 35.

Lopez Bernal, J. et al. Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594 (2021).

36. 36.

*Investigation of Novel SARS-CoV-2 Variants of Concern* Technical briefing 19  
(Public Health England, 2021);  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1005517/Technical\\_Briefing\\_19.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005517/Technical_Briefing_19.pdf)

37. 37.

Ferreira, I. et al. SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.08.443253> (2021).

38. 38.

Wall, E. C. et al. Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination. *Lancet* **397**, 2331–2333 (2021).

39. 39.

Steel, K. & Haughton, P. *Coronavirus (COVID-19) Infection Survey, Antibody and Vaccination Data, UK* (ONS, 2021);  
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionssurveyantibodyandvaccinationdatafortheuk/21july2021>

40. 40.

Anderson, R. et al. *Reproduction Number ( $R$ ) and Growth Rate ( $r$ ) of the COVID-19 Epidemic in the UK: Methods of Estimation, Data Sources, Causes of Heterogeneity, and use as a Guide in Policy Formulation* (The Royal Society, 2020).

41. 41.

Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846–849 (2020).

42. 42.

Funk, S. et al. Combining serological and contact data to derive target immunity levels for achieving and maintaining measles elimination. *BMC Med.* **17**, 180 (2019).

43. 43.

Hodgson, D., Flasche, S., Jit, M., Kucharski, A. J. & CMMID COVID-19 Working Group The potential for vaccination-induced herd immunity against the SARS-CoV-2 B.1.1.7 variant. *Euro Surveill.* **26**, 2100428 (2021).

44. 44.

van Dorp, L., Houldcroft, C. J., Richard, D. & Balloux, F. COVID-19, the first pandemic in the post-genomic era. *Curr. Opin. Virol.* **50**, 40–48 (2021).

45. 45.

The Health Protection (Coronavirus, Restrictions) (England) (No. 4) Regulations 2020. *Wikipedia* [https://en.wikipedia.org/w/index.php?title=The\\_Health\\_Protection\\_\(Coronavirus,\\_Restrictions\)\\_\\_\(England\)\\_\(\\_No.\\_4\)\\_Regulations\\_2020&oldid=1014701607](https://en.wikipedia.org/w/index.php?title=The_Health_Protection_(Coronavirus,_Restrictions)__(England)_(_No._4)_Regulations_2020&oldid=1014701607) (2021).

46. 46.

Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).

47. 47.

Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).

48. 48.

Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).

49. 49.

Bingham, E. et al. Pyro: deep universal probabilistic programming. Preprint at <http://arxiv.org/abs/1810.09538> (2018).

50. 50.

Phan, D., Pradhan, N. & Jankowiak, M. Composable effects for flexible and accelerated probabilistic programming in NumPyro. Preprint at <http://arxiv.org/abs/1912.11554> (2019).

51. 51.

Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. Stochastic variational inference. *J. Mach. Learn. Res.* **14**, 1303–1347 (2013).

52. 52.

Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

53. 53.

Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86 (1986).

54. 54.

Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).

55. 55.

De Maio, N., Wu, C.-H., O'Reilly, K. M. & Wilson, D. New routes to phylogeography: a bayesian structured coalescent approximation. *PLoS Genet.* **11**, e1005421 (2015).

56. 56.

Turakhia, Y. et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat. Genet.* **53**, 809–816 (2021).

57. 57.

Saguleko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).

58. 58.

Gill, M. S. et al. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).

59. 59.

Sherratt, K. et al. Exploring surveillance data biases when estimating the reproduction number: with insights into subpopulation transmission of COVID-19 in England. *Philos. Trans. Royal Soc. B* **376**  
<https://doi.org/10.1098/RSTB.2020.0283> (2021).

60. 60.

Abbott, S. et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res.* **5**, 112 (2020).

61. 61.

Hellewell, J. et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Med.* **19**, 106 (2021).

62. 62.

Hart, W. S. et al. Inference of SARS-CoV-2 generation times using UK household data. Preprint at *medRxiv* <https://doi.org/10.1101/2021.05.27.21257936> (2021).

63. 63.

Pouwels, K. B. et al. Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey. *Lancet Publ. Health* **6**, e30–e38 (2021).

64. 64.

Donnarumma, K. S. *Coronavirus (COVID-19) Infection Survey, UK* (ONS, 2021);  
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionssurveypilot/23april2021>

## Acknowledgements

We thank E. Allara (Cambridge) and G. Whitton (Sanger) for providing outer postcodes to LTLA mappings; R. Beale for comments and J. McCrone for setting up Thorney Beast analysis; all of the contributors who submitted genome sequences to GISAID (acknowledgement tables for individual sequences are provided at GitHub; <https://github.com/NicolaDM/phylogeographySARS-CoV-2>); and our colleagues at EMBL-EBI, the Wellcome Sanger Institute and COG-UK for discussions and comments on this manuscript. COG-UK is supported by funding from the Medical Research Council (MRC), part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. Additional sequence generation was funded by the Department of Health and Social Care. H.S.V., J.P.G. and M.G. are supported by a grant from the Department of Health and Social Care. A.W.J., E.B. and M.G. are beneficiaries from grant NNF17OC0027594 from the Novo Nordisk Foundation. E.V. is supported by Wellcome Trust grant 220885/Z/20/Z. T.S. is supported by grant 210918/Z/18/Z, and J.H. and S.F. by grant 210758/Z/18/Z from the Wellcome Trust. H.S.V., N.D.M., A.W.J., N.G., E.B. and M.G. are supported by EMBL.

## Funding

Open access funding provided by Deutsches Krebsforschungszentrum (DKFZ).

## Author information

## Affiliations

1. European Molecular Biology Laboratory, European Bioinformatics Institute  
EMBL-EBI, Hinxton, UK

Harald S. Vöhringer, Nicola De Maio, Alexander W. Jung, Nick Goldman, Ewan Birney & Moritz Gerstung

2. Wellcome Sanger Institute, Hinxton, UK

Theo Sanderson, Matthew Sinnott, Thuy Nguyen, Richard Goater, Frank Schwach, Cristina V. Ariani, Sonia Gonçalves, David K. Jackson, Ian Johnston, Callum Saint, John Sillitoe, Maria Suciu, Irina Abnizova, Louise Aigrain, Alex Alderton, Mozam Ali, Laura Allen, Roberto Amato, Ralph Anderson, Cristina Ariani, Siobhan Austin-Guest, Sendu Bala, Jeffrey Barrett, Andrew Bassett, Kristina Battleday, James Beal, Mathew Beale, Charlotte Beaver, Sam Bellany, Tristram Bellerby, Katie Bellis, Duncan Berger, Matt Berriman, Emma Betteridge, Paul Bevan, Simon Binley, Jason Bishop, Kirsty Blackburn, James Bonfield, Nick Boughton, Sam Bowker, Timothy Brendler-Spaeth, Iraad Bronner, Tanya Brooklyn, Sarah Kay Buddenborg, Robert Bush, Catarina Caetano, Alex Cagan, Nicola Carter, Joanna Cartwright, Tiago Carvalho Monteiro, Liz Chapman, Tracey-Jane Chillingworth, Peter Clapham, Richard Clark, Adrian Clarke, Catriona Clarke, Daryl Cole, Elizabeth Cook, Maria Coppola, Linda Cornell, Clare Cornwell, Craig Corton, Abby Crackett, Alison Cranage, Harriet Craven, Sarah Craw, Mark Crawford, Tim Cutts, Monika Dabrowska, Matt Davies, Robert Davies, Joseph Dawson, Callum Day, Aiden Densem, Thomas Dibling, Cat Dockree, David Dodd, Sunil Dogga, Matthew Dorman, Gordon Dougan, Martin Dougherty, Alexander Dove, Lucy Drummond, Eleanor Drury, Monika Dudek, Jillian Durham, Laura Durrant, Elizabeth Easthope, Sabine Eckert, Pete Ellis, Ben Farr, Michael Fenton, Marcella Ferrero, Neil Flack, Howard Fordham, Grace Forsythe, Luke Foulser, Matt Francis, Audrey Fraser, Adam Freeman, Anastasia Galvin, Maria Garcia-Casado, Alex Gedny, Sophia Girgis, James Glover, Sonia Goncalves, Scott Goodwin, Oliver Gould, Marina Gourtovaia, Andy Gray, Emma Gray, Coline Griffiths, Yong Gu, Florence Guerin, Will Hamilton, Hannah Hanks, Ewan Harrison, Alexandria Harrott, Edward Harry, Julia Harvison, Paul Heath, Anastasia Hernandez-Koutoucheva, Rhiannon Hobbs, Dave Holland, Sarah Holmes, Gary Hornett, Nicholas Hough, Liz Huckle, Lena Hughes-Hallet, Adam Hunter, Stephen Inglis, Sameena Iqbal, Adam Jackson, David Jackson, Keith James, Dorota Jamrozy, Carlos Jimenez Verdejo, Matthew Jones, Kalyan Kallepally, Leanne Kane, Keely Kay, Sally Kay, Jon Keatley, Alan Keith, Alison

King, Lucy Kitchin, Matt Kleanthous, Martina Klimekova, Petra Korlevic, Ksenia Krasheninnikova, Greg Lane, Cordelia Langford, Adam Laverack, Katharine Law, Mara Lawniczak, Stefanie Lensing, Steven Leonard, Laura Letchford, Kevin Lewis, Amanah Lewis-Wade, Jennifer Liddle, Quan Lin, Sarah Lindsay, Sally Linsdell, Rich Livett, Stephanie Lo, Rhona Long, Jamie Lovell, Jon Lovell, Catherine Ludden, James Mack, Mark Maddison, Aleksei Makunin, Irfan Mamun, Jenny Mansfield, Neil Marriott, Matt Martin, Matthew Mayho, Shane McCarthy, Jo McClintock, Samantha McGuigan, Sandra McHugh, Liz McMinn, Carl Meadows, Emily Mobley, Robin Moll, Maria Morra, Leanne Morrow, Kathryn Murie, Sian Nash, Claire Nathwani, Plamena Naydenova, Alexandra Neaverson, Rachel Nelson, Ed Nerou, Jon Nicholson, Tabea Nimz, Guillaume G. Noell, Sarah O'Meara, Valeriu Ohan, Karen Oliver, Charles Olney, Doug Ormond, Agnes Oszlanczi, Steve Palmer, Yoke Fei Pang, Barbora Pardubska, Naomi Park, Aaron Parmar, Gaurang Patel, Minal Patel, Maggie Payne, Arabella Petersen, Deborah Plowman, Tom Preston, Liam Prestwood, Christoph Puethe, Michael Quail, Diana Rajan, Shavanti Rajatileka, Richard Rance, Suzannah Rawlings, Nicholas Redshaw, Joe Reynolds, Mark Reynolds, Simon Rice, Matt Richardson, Connor Roberts, Katrina Robinson, Melanie Robinson, David Robinson, Hazel Rogers, Eduardo Martin Rojo, Daljit Roopra, Mark Rose, Luke Rudd, Nicholas Salmon, David Saul, Frank Schwach, Carol Scott, Phil Seekings, Lesley Shirley, Alison Simms, Matthew Sinnott, Shanthi Sivadasan, Bart Siwek, Dale Sizer, Kenneth Skeldon, Jason Skelton, Joanna Slater-Tunstill, Lisa Sloper, Nathalie Smerdon, Chris Smith, Christen Smith, James Smith, Katie Smith, Michelle Smith, Sean Smith, Tina Smith, Leighton Sneade, Carmen Diaz Soria, Emily Souster, Andrew Sparkes, Michael Spencer-Chapman, Janet Squares, Robert Stanley, Claire Steed, Tim Stickland, Ian Still, Michael R. Stratton, Michelle Strickland, Allen Swann, Agnieszka Swiatkowska, Neil Sycamore, Emma Swift, Edward Symons, Suzanne Szluha, Emma Taluy, Nunu Tao, Katy Taylor, Sam Taylor, Stacey Thompson, Mark Thompson, Mark Thomson, Nicholas Thomson, Scott Thurston, Gerry Tonkin-Hill, Dee Toombs, Benjamin Topping, Jaime Tovar-Corona, Daniel Ungureanu, James Uphill, Jana Urbanova, Philip Jansen Van Vuuren, Valerie Vancollie, Paul Voak, Danielle Walker, Matthew Walker, Matt Waller, Gary Ward, Charlie Weatherhogg, Niki Webb, Danni Weldon, Alan Wells, Eloise Wells, Luke Westwood, Theo Whipp, Thomas Whiteley, Georgia Whitton, Andrew Whitwham, Sara Widaa, Mia Williams, Mark Wilson, Sean Wright, Ewan M. Harrison, Sónia Gonçalves, David M. Aanensen, Dinesh Aggarwal, Anthony P. Underwood, Cordelia F. Langford, Dominic Kwiatkowski, Andrew R. Bassett, Michael H. Spencer Chapman, Theo Sanderson, Naomi R. Park, Robert M. Davies, Katherine L. Bellis, Matthew J. Dorman, Robin J. Moll, Jaime M.

Tovar-Corona, Khalil Abudahab, Ben E. W. Taylor, Jon-Paul Keatley, Iraad F. Bronner, Ben W. Farr, Stefanie V. Lensing, Shane A. McCarthy, Michael A. Quail, Nicholas M. Redshaw, Scott A. J. Thurston, Jennifer Liddle, Dominic Kwiatkowski, Inigo Martincorena & Jeffrey C. Barrett

3. The Francis Crick Institute, London, UK

Theo Sanderson

4. Public Health England, London, UK

Frank Schwach, Ian Harrison, Catherine Ludden, Dinesh Aggarwal, Husam Osman, Meera Chand, Sharon J. Peacock, Peter Muir, Esther Robinson, Barry B. Vipond, Andrew Bosworth, Stephanie Hutchings, Hannah M. Pymont, Richard Myers, Alicia Thornton, Katerina Galai, Shazaad S. Y. Ahmad, Nicholas W. Machin, Richard Hopes, Chloe Bishop, Vicki Chalker, Elias Allara, Clare Pearson, David Bibby, Gavin Dabrera, Nicholas Ellaby, Eileen Gallagher, Jonathan Hubb, Angie Lackenby, David Lee, Nikos Manesis, Tamyo Mbisa, Steven Platt, Katherine A. Twohig & Meera Chand

5. London School of Hygiene & Tropical Medicine, London, UK

Joel Hellewell & Sebastian Funk

6. The Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Jasmina Panovska-Griffiths, Tanya Golubchik, David Bonsall, Katrina Lythgoe, Helen Fryer, Christophe Fraser & Laura Thomson

7. MRC Centre for Global Infectious Disease Analysis, Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK

Adam A. Witney, Erik M. Volz, Rob Johnson, Olivia Boyd, Lily Geidelberg, Manon Ragonnet-Cronin & Erik Volz

8. Guy's and St Thomas' NHS Foundation Trust, London, UK

Perminder Smith, Adela Alcolea, Natasha Ohemeng-Kumi, John Ramble, Jasveen Sehmi & Meera Chand

9. Division for AI in Oncology, German Cancer Research Centre DKFZ, Heidelberg, Germany

Moritz Gerstung

10. Department of Medicine, University of Cambridge, Cambridge, UK

Sharon Peacock, Ramin Sadri, Catarina Sousa, Catherine Ludden, M. Estee Torok, Dinesh Aggarwal, Alessandro M. Carabelli, Sharon J. Peacock, Beth Blane, MacGregor Cox, Patrick Maxwell, Ken Smith, Ellena Brooks, Carol M. Churcher, Mireille Fragakis, Katerina Galai, Andrew Jermy, Sarah Judges, Georgina M. McManus, Kim S. Smith, Elaine Westwick & Theresa Feltwell

11. Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, UK

Samuel C. Robson, Angela H. Beckett & Ekaterina Shelest

12. School of Pharmacy & Biomedical Sciences, University of Portsmouth, Portsmouth, UK

Samuel C. Robson, Katie F. Loveson, Kate F. Cook, Christopher Fearn & Salman Goudarzi

13. Cardiff University, Cardiff, UK

Thomas R. Connor, Anna Price, Andrew Mack, Arthur Morriss, Thomas Whalley, Joel Southgate, Martyn Guest, Christine Kitchen, Ian Merrick, Robert Munn, Owen Jones, Catherine Bresner, William Fuller, Angela Marchbank & Trudy Workman

14. Public Health Wales, Cardiff, UK

Thomas R. Connor, Sally Corden, Catherine Moore, Joanne Watkins, Matthew Bull, Catryn Williams, Nicole Pacchiarini, Simon Cottrell, Sara Rey, David Heyburn, Chris Williams, Noel Craine, Malorie Perry, Alexander Adams, Tara Annett, Hibo Asad, Alec Birchley, Jason Coombes, Johnathan M. Evans, Laia Fina, Bree Gatica-Wilcox, Lauren Gilbert, Lee Graham, Jessica Hey, Ember Hilvers, Sophie Jones, Hannah Jones, Sara Kumzien-Summerhayes, Caoimhe McKerr, Jessica Powell, Georgia Pugh, Sarah Taylor, Joel Southgate, Alisha Davies, Elen De Lacy, Fatima Downing, Sue Edwards, Laura Gifford, Mari Morgan & Amy Gaskin

15. Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK

Nicholas J. Loman, Sam Nicholls, Claire McMurray, Alan McNally, Joshua Quick, Radoslaw Poplawski, Joanne Stockton & Will Rowe

16. King's College London, London, UK

Rocio T. Martinez Nunez & Themoula Charalampous

17. University of Edinburgh, Edinburgh, UK

Andrew Rambaut, Kate E. Templeton, Áine O'Toole, Thomas Williams, Rachel Colquhoun, Verity Hill, Ben Jackson, J. T. McCrone, Nathan Medd, Emily Scher, Carlos E. Balcazar, Michael D. Gallagher, Daniel Maloney, Thomas D. Stanton & Kathleen A. Williamson

18. Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK

Luke B. Snell, Gaia Nebbia, Jonathan Edgeworth, Rahul Batra, Bindi Patel, Themoula Charalampous & Amita Patel

19. University College London Hospitals NHS Foundation Trust, London, UK

Eleni Nastouli, Catherine Houlihan, Matthew Byott, Sunando Roy, Judith Heaney, Judith Breuer, Moira J. Spyer, Rachel J. Williams, Charlotte A. Williams, John A. Hartley, Leah Ensell, Helen L. Lowe, Laurentiu Maftei, Matteo Mondani, Nadua Bayzid & Marius Cotic

20. Advanced Pathogen Diagnostics Unit, University College London Hospital, London, UK

Eleni Nastouli, Dan Frampton, Matthew Byott, Judith Heaney & Moira J. Spyer

21. Great Ormond Street Institute of Child Health, University College London, London, UK

Eleni Nastouli & Moira J. Spyer

22. Department of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

M. Estee Torok & William L. Hamilton

23. Division of Virology, Department of Pathology, University of Cambridge, Cambridge, UK

Ian G. Goodfellow, Yasmin Chaudhry, Rhys Izuagbe, Aminu S. Jahun, Iliana Georgana, Myra Hosmillo & Malte L. Pinckert

24. University Hospital Southampton NHS Foundation Trust, Southampton, UK

Jacqui A. Prieto, Kordo Saeed, Stephen Aplin, Matthew Harvey, Thea Sass, Helen Umpleby, Helen Wheeler, Emanuela Pelosi, Siona Silviera, Eleri Wilson-Davies, Adhyana I. K. Mahanama, Buddhini Samaraweera & Sarah Jeremiah

25. School of Health Sciences, University of Southampton, Southampton, UK

Jacqui A. Prieto

26. School of Medicine, University of Southampton, Southampton, UK

Kordo Saeed

27. Division of Infection and Immunity, University College London, London, UK

Catherine Houlihan & Dan Frampton

28. University of Brighton, Brighton, UK

Giselda Bucca, Colin P. Smith & Andrew R. Hesketh

29. Institute for Infection and Immunity, St George's University of London, London, UK

Cassie F. Pope, Kenneth G. Laing & Irene M. Monahan

30. Infection Care Group, St George's University Hospitals NHS Foundation Trust, London, UK

Cassie F. Pope

31. MRC–University of Glasgow Centre for Virus Research, Glasgow, UK

Emma C. Thomson, James G. Shepherd, David L. Robertson, William T. Harvey, Joseph Hughes, Richard J. Orton, Ana da Silva Filipe, Igor Starinskij, Sreenu Vattipally, Afrida Mukaddas, Derek W. Wright, Alice Broos, Daniel Mair, Jenna Nichols, Kyriaki Nomikou, Lily Tong & Ioulia Tsatsani

32. University of Cambridge, Cambridge, UK

Ewan M. Harrison, Ravi K. Gupta, Michael H. Spencer Chapman, Chris Ruis, Sophia T. Girgis, Katherine L. Bellis, Leanne M. Kermack, Claire Cormie, Joana Dias, Sally Forrest, Ellen E. Higginson, Mailis Maes, Jamie Young, Elias Allara, Clare Pearson & Shane A. McCarthy

33. Queen's University Belfast, Belfast, UK

Fiona Rogan, Derek J. Fairley, David A. Simpson, Marc Fuchs, Julia Miskelly, Stephen Bridgett, Timofey Skvortsov, Declan T. Bradley, Zoltan Molnar, Chris Baxter, Sílvia F. Carvalho, Deborah Lavin, Arun Mariappan, Clara Radulescu, Aditi Singh & Miao Tang

34. Blackpool Teaching Hospitals NHS Foundation Trust, Blackpool, UK

Shaun M. Beckwith, Abigail Murray & Dawn Singleton

35. Hull University Teaching Hospitals NHS Trust, Hull, UK

Kirstine Eastick, Patrick J. Lillie, Phillipa J. Burns, Kavitha Gajee, Katie Kitchman & William Everson

36. University Hospitals Dorset NHS Foundation Trust, Poole, UK

Liz A. Sheridan, Dorian Crudginton & Ben Macklin

37. University Hospitals Sussex NHS Foundation Trust, Worthing, UK

Paul Randell, Mohammed O. Hassan-Ibrahim, Michelle J. Erkert, Nicola J. Chaloner, Benjamin J. Cogger, Lisa J. Easton, Hannah Huckson, Jonathan Lewis, Sarah Lowdon, Cassandra S. Malone, Florence Munemo, Manasa Mutingwende, Roberto Nicodemi, Olga Podplomyk & Thomas Somassa

38. University of Exeter, Exeter, UK

Leigh M. Jackson, Ben Temperton, Stephen L. Michell, Aaron R. Jeffries, Christopher R. Jones, Bridget A. Knight, Robin Manley, Michelle L. Michelsen, Christine M. Sambles, David J. Studholme & Joanna Warwick-Dugdale

39. Belfast Health & Social Care Trust, Belfast, UK

Derek J. Fairley, James P. McKenna & Tanya Curran

40. Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

Matthew W. Loose, Matthew Carlile, Nadine Holmes, Christopher Moore, Victoria Wright, Johnny Debebe & Fei Sang

41. East Kent Hospitals University NHS Foundation Trust, Canterbury, UK

Samuel Moses & Hannah Lowe

42. University of Kent, Canterbury, UK

Samuel Moses

43. Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Matthew Bashton & Gregory R. Young

44. Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Matthew Bashton, Gregory R. Young, Clare M. McCann, Andrew Nelson, Matthew R. Crown, John H. Henderson, Amy Hollis, William Stanley & Wen C. Yew

45. NU-OMICS, Northumbria University, Newcastle upon Tyne, UK

Darren L. Smith, Clare M. McCann & Andrew Nelson

46. Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, UK

David M. Aanensen, Anthony P. Underwood, Khalil Abudahab, Mirko Menegazzo, Ben E. W. Taylor & Corin A. Yeats

47. Public Health England, Cambridge, Cambridge, UK

Martin D. Curran & Surendra Parmar

48. University of Sheffield, Sheffield, UK

Matthew D. Parker, Thushan I. de Silva, Dennis Wang, Mehmet Yavus, Samantha E. Hansford, Benjamin H. Foulkes, Marta Gallis, Hailey R. Hornsby, Stavroula F. Louka, Manoj Pohare, Paige Wolverson, Peijun Zhang, Timothy M. Freeman, Sharon N. Hsu, Benjamin B. Lindsey, Nikki Smith, Cariad Evans, Kate

Johnson, David G. Partridge, Mohammad Raza, Alexander J. Keeley, Adrienn Angyal, Luke R. Green & Max Whiteley

49. Portsmouth Hospitals University NHS Trust, Portsmouth, UK

Sharon Glaysher, Scott Elliott, Kelly Bicknell, Robert Impey & Sarah Wyllie

50. NHS Lothian, Edinburgh, UK

Kate E. Templeton, Rebecca Dewar, Martin P. McHugh, Elizabeth Wastnedge, Seb Cotton & Abbie Gallagher

51. NHS Greater Glasgow and Clyde, Glasgow, UK

Rory N. Gunson, Rachel Blacow & Jon Perkins

52. Quadram Institute Bioscience, Norwich, UK

Justin O'Grady, Andrew J. Page, Gemma L. Kay, Nabil-Fareed Alikhan, Alexander J. Trotter, Leonardo de Oliveira Martins, Alison E. Mather, Lizzie Meadows, Alp Aydin, David J. Baker, Ebenezer Foster-Nyarko, Sophie J. Prosolek, Steven Rudder & Thanh Le-Viet

53. University of East Anglia, Norwich, UK

Justin O'Grady & Rose K. Davidson

54. University of Oxford, Oxford, UK

Dominic Kwiatkowski

55. Hampshire Hospitals NHS Foundation Trust, Basingstoke, UK

Nicholas Cortes, Nathan Moore, Claire Thomas, Stephen P. Kidd, Beatrice Bertolusso, Jessica Lynch, Gabrielle Vernet & Rebecca Williams

56. Royal Free London NHS Foundation Trust, London, UK

Tabitha W. Mahungu, Tanzina Haque, Jennifer Hart, Dianne Irish-Tavares & Eric Witele

57. South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK

Steven Liggett, Craig Mower, Louisa K. Watson, Paul Baker, Stephen Bonner, Sarah Essex & Leanne J. Murray

58. School of Biological Sciences, University of Portsmouth, Portsmouth, UK

Angela H. Beckett, Yann Bourgeois & Ekaterina Shelest

59. Public Health Scotland, Edinburgh, UK

Matthew T. G. Holden, Sharif Shaaban, Stefan Rooke & Gonzalo Yebra

60. Health Services Laboratories, London, UK

Lisa J. Levett, Judith Heaney, Paul R. Grant, Stuart Kirk, Wendy Chatterton & Monika Pusok

61. Heartlands Hospital, Birmingham, UK

Husam Osman, Esther Robinson, Li Xu-McCrae & Andrew Bosworth

62. University of Liverpool, Liverpool, UK

Alistair C. Darby, Steve Paterson, Miren Iturriza-Gomara, Anita O. Lucaci, Sam T. Haldenby, Kathryn A. Jackson, Lance Turtle, Richard Eccles, Matthew Gemmell, Richard Gregory, Margaret Hughes, Charlotte Nelson, Lucille Rainbow, Edith E. Vamos, Hermione J. Webster, Mark Whitehead & Claudia Wierzbicki

63. Department of Zoology, University of Oxford, Oxford, UK

Oliver G. Pybus, Louis du Plessis, Moritz U. G. Kraemer, Jayna Raghwani, Alex E. Zarebski, Stephen W. Attwood, Marina Escalera Zamudio, Sarah Francois, Bernardo Gutierrez & Tetyana I. Vasylyeva

64. MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Daniela de Angelis, Chris J. Illingworth, Chris Jackson & David Pascall

65. Microbiology Department, Buckinghamshire Healthcare NHS Trust, Aylesbury, UK

Nick Wong, David T. Pritchard, Debbie Binns & Victoria James

66. The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

Yusri Taha, Jennifer Collins, Gary Eltringham, Shirelle Burton-Fanning, Brendan A. I. Payne & Sheila Waugh

67. University of St Andrews, St Andrews, UK

Martin P. McHugh

68. Imperial College Healthcare NHS Trust, London, UK

Siddharth Mookerjee, Alison Cox, Pinglawathee Madona, Marcus Pond, Paul A. Randell, Alison H. Holmes, James R. Price & Frances Bolt

69. NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, London, UK

Siddharth Mookerjee, Alison H. Holmes, James R. Price & Frances Bolt

70. Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Ben Warne

71. Department of Microbiology, South West London Pathology, London, UK

Joshua F. Taylor & Ngee Keong Tan

72. Betsi Cadwaladr University Health Board, Wrexham, UK

Helen Adams

73. Institute of Biodiversity, Animal Health & Comparative Medicine, Glasgow, UK

William T. Harvey & Katherine L. Smollett

74. Norfolk County Council, Norfolk, UK

Lewis G. Spurgin & Louise Smith

75. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Judith Breuer, Julianne R. Brown, Kathryn A. Harris, Nathaniel Storey, Laura Atkinson, Jack C. D. Lee & Divya Shah

76. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

John A. Todd, David Buck, Angie Green, George MacIntyre-Cockett & Amy Trebes

77. Cardiff and Vale University Health Board, Cardiff, UK

Michaela John, Sian Morgan, Safiah Afifi, Robert Beer, Joshua Maksimovic, Kathryn McCluggage Masters & Karla Spellman

78. Turnkey Laboratory, University of Birmingham, Birmingham, UK

Alan McNally, Fiona Ashford, Angus Best, Liam Crawford, Nicola Cumley, Megan Mayhew, Oliver Megram, Jeremy Mirza, Emma Moles-Garcia & Benita Percival

79. Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK

Samir Dervisevic, Rachael Stanley & Lindsay Coupland

80. Royal Brompton and Harefield Hospitals, London, UK

Newara A. Ramadan

81. The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, King's Lynn, UK

Christopher Jeanes

82. Whittington Health NHS Trust, London, UK

Jana Catalan & Neil Jones

83. Barking, Havering and Redbridge University Hospitals NHS Trust, London, UK

Amy Ash & Cherian Koshy

84. Bournemouth University, Bournemouth, UK

Magdalena Barrow, Sarah L. Buchan & Anna Mantzouratou

85. Clinical Microbiology Department, Queens Medical Centre, Nottingham University Hospitals NHS Trust, Nottingham, UK

Gemma Clark, Louise Berry, Tim Boswell, Vicki M. Fleming, Hannah C. Howson-Wells, Amelia Joseph, Manjinder Khakh & Michelle M. Lister

86. Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, UK

Christopher W. Holmes, Paul W. Bird, Karlie Fallon, Thomas Helmer, Claire L. McMurray, Mina Odedra, Jessica Shaw, Julian W. Tang & Nicholas J. Willford

87. County Durham and Darlington NHS Foundation Trust, Durham, UK

Sharon Campbell, Victoria Blakey, Veena Raviprakash, Nicola Sheriff & Lesley-Anne Williams

88. Department of Microbiology, Kettering General Hospital, Kettering, UK

Thomas Davis, Sahar Eldirdiri & Anita Kenyon

89. Barts Health NHS Trust, London, UK

Kathryn A. Harris

90. North West London Pathology, London, UK

Alison Cox, Pinglawathee Madona, Marcus Pond & Paul A. Randell

91. Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK

Karen T. Withell & Graciela Sluga

92. Microbiology, Royal Oldham Hospital, Oldham, UK

Cheryl Williams

93. North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK

Clive Graham, Edward Barton, Debra Padgett & Garren Scott

94. North Tees and Hartlepool NHS Foundation Trust, Stockton on Tees, UK

Rebecca Denton-Smith, Emma Swindells, Robyn Turnbull & Jane Greenaway

95. Path Links, Northern Lincolnshire and Goole NHS Foundation Trust, Grimsby, UK

Tim J. Sloan, Phillip Clarke, Nichola Duckworth & Sarah Walsh

96. Queen Elizabeth Hospital, Birmingham, UK

Anna Casey & Liz Ratcliffe

97. Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Christopher R. Jones, Bridget A. Knight, Jennifer Poyner, Cressida Auckland & Helen Morcrette

98. Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK

Patrick C. McClure, Jonathan Ball, Timothy Byaruhanga, Joseph G. Chappell, Jayasree Dey, Jack D. Hill & Emily J. Park

99. Swansea University, Swansea, UK

Ronan A. Lyons

100. Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, London, UK

Perminder Smith, Adela Alcolea, Natasha Ohemeng-Kumi, John Ramble & Jasveen Sehmi

101. Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Mehmet Yavus, Cariad Evans, Kate Johnson, David G. Partridge & Mohammad Raza

102. Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

Nicola Reynolds

103. West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, UK

Lynne Ferguson, Emily J. Goldstein, Alasdair Maclean & Rachael Tomb

104. Public Health Agency, Belfast, UK

Tim Wyatt & Declan T. Bradley

105. Northumbria Healthcare NHS Foundation Trust, Newcastle upon Tyne, UK

Giles Idle & Kevin Cole

106. University of Southampton, Southampton, UK

Matilde Mori

107. East Suffolk and North Essex NHS Foundation Trust, Colchester, UK

Luke Bedford

108. East Sussex Healthcare NHS Trust, St Leonards-on-Sea, UK

James S. Cargill & Warwick Hughes

109. Gateshead Health NHS Foundation Trust, Gateshead, UK

Jonathan Moore & Susanne Stonehouse

110. Isle of Wight NHS Trust, Newport, UK

Anibolina Castigador & Emily Macnaughton

111. King's College Hospital NHS Foundation Trust, London, UK

Kate El Bouzidi, Temi Lampejo & Malur Sudhanva

112. Liverpool Clinical Laboratories, Liverpool, UK

Cassie Breen

113. Manchester University NHS Foundation Trust, Manchester, UK

Shazaad S. Y. Ahmad, Ryan P. George & Nicholas W. Machin

114. North Middlesex University Hospital NHS Trust, London, UK

Aidan Cross & Mariyam Mirfenderesky

115. Southwest Pathology Services, Taunton, UK

Andrew I. Lawton

116. The Royal Marsden NHS Foundation Trust, London, UK

Andrea N. Gomes, Maimuna Kimuli & Darren R. Murray

117. The Royal Wolverhampton NHS Trust, Wolverhampton, UK

Paula Ashfield & Donald Dobie

118. University of Birmingham, Birmingham, UK

Andrew Beggs & Alex Richter

119. Watford General Hospital, Watford, UK

Arezou Fanaie, Rachel A. Hilson & Geraldine Yaze

120. Guy's and St Thomas' Biomedical Research Centre, London, UK

Flavia Flaviani

121. Newcastle University, Newcastle, UK

Sarah O'Brien, Steven Rushton & Roy Sanderson

## Consortia

### The Wellcome Sanger Institute COVID-19 Surveillance Team

- Irina Abnizova
- , Louise Aigrain
- , Alex Alderton
- , Mozam Ali
- , Laura Allen
- , Roberto Amato
- , Ralph Anderson
- , Cristina Ariani
- , Siobhan Austin-Guest
- , Sendu Bala
- , Jeffrey Barrett
- , Andrew Bassett
- , Kristina Battleday
- , James Beal
- , Mathew Beale
- , Charlotte Beaver
- , Sam Bellany
- , Tristram Bellerby
- , Katie Bellis
- , Duncan Berger
- , Matt Berriman

- , Emma Betteridge
- , Paul Bevan
- , Simon Binley
- , Jason Bishop
- , Kirsty Blackburn
- , James Bonfield
- , Nick Boughton
- , Sam Bowker
- , Timothy Brendler-Spaeth
- , Iraad Bronner
- , Tanya Brooklyn
- , Sarah Kay Buddenborg
- , Robert Bush
- , Catarina Caetano
- , Alex Cagan
- , Nicola Carter
- , Joanna Cartwright
- , Tiago Carvalho Monteiro
- , Liz Chapman
- , Tracey-Jane Chillingworth
- , Peter Clapham
- , Richard Clark
- , Adrian Clarke
- , Catriona Clarke
- , Daryl Cole
- , Elizabeth Cook
- , Maria Coppola
- , Linda Cornell
- , Clare Cornwell
- , Craig Corton
- , Abby Crackett
- , Alison Cranage
- , Harriet Craven
- , Sarah Craw
- , Mark Crawford
- , Tim Cutts
- , Monika Dabrowska
- , Matt Davies
- , Robert Davies
- , Joseph Dawson
- , Callum Day
- , Aiden Densem

- , Thomas Dibling
- , Cat Dockree
- , David Dodd
- , Sunil Dogga
- , Matthew Dorman
- , Gordon Dougan
- , Martin Dougherty
- , Alexander Dove
- , Lucy Drummond
- , Eleanor Drury
- , Monika Dudek
- , Jillian Durham
- , Laura Durrant
- , Elizabeth Easthope
- , Sabine Eckert
- , Pete Ellis
- , Ben Farr
- , Michael Fenton
- , Marcella Ferrero
- , Neil Flack
- , Howerd Fordham
- , Grace Forsythe
- , Luke Foulser
- , Matt Francis
- , Audrey Fraser
- , Adam Freeman
- , Anastasia Galvin
- , Maria Garcia-Casado
- , Alex Gedny
- , Sophia Girgis
- , James Glover
- , Sonia Goncalves
- , Scott Goodwin
- , Oliver Gould
- , Marina Gourtovaia
- , Andy Gray
- , Emma Gray
- , Coline Griffiths
- , Yong Gu
- , Florence Guerin
- , Will Hamilton
- , Hannah Hanks

- , Ewan Harrison
- , Alexandria Harrott
- , Edward Harry
- , Julia Harvison
- , Paul Heath
- , Anastasia Hernandez-Koutoucheva
- , Rhianon Hobbs
- , Dave Holland
- , Sarah Holmes
- , Gary Hornett
- , Nicholas Hough
- , Liz Huckle
- , Lena Hughes-Hallet
- , Adam Hunter
- , Stephen Inglis
- , Sameena Iqbal
- , Adam Jackson
- , David Jackson
- , Keith James
- , Dorota Jamrozy
- , Carlos Jimenez Verdejo
- , Ian Johnston
- , Matthew Jones
- , Kalyan Kallepally
- , Leanne Kane
- , Keely Kay
- , Sally Kay
- , Jon Keatley
- , Alan Keith
- , Alison King
- , Lucy Kitchin
- , Matt Kleanthous
- , Martina Klimekova
- , Petra Korlevic
- , Ksenia Krasheninnkova
- , Dominic Kwiatkowski
- , Greg Lane
- , Cordelia Langford
- , Adam Laverack
- , Katharine Law
- , Mara Lawniczak
- , Stefanie Lensing

- , Steven Leonard
- , Laura Letchford
- , Kevin Lewis
- , Amanah Lewis-Wade
- , Jennifer Liddle
- , Quan Lin
- , Sarah Lindsay
- , Sally Linsdell
- , Rich Livett
- , Stephanie Lo
- , Rhona Long
- , Jamie Lovell
- , Jon Lovell
- , Catherine Ludden
- , James Mack
- , Mark Maddison
- , Aleksei Makunin
- , Irfan Mamun
- , Jenny Mansfield
- , Neil Marriott
- , Matt Martin
- , Inigo Martincorena
- , Matthew Mayho
- , Shane McCarthy
- , Jo McClintock
- , Samantha McGuigan
- , Sandra McHugh
- , Liz McMinn
- , Carl Meadows
- , Emily Mobley
- , Robin Moll
- , Maria Morra
- , Leanne Morrow
- , Kathryn Murie
- , Sian Nash
- , Claire Nathwani
- , Plamena Naydenova
- , Alexandra Neaverson
- , Rachel Nelson
- , Ed Nerou
- , Jon Nicholson
- , Tabea Nimz

- , Guillaume G. Noell
- , Sarah O'Meara
- , Valeriu Ohan
- , Karen Oliver
- , Charles Olney
- , Doug Ormond
- , Agnes Oszlanczi
- , Steve Palmer
- , Yoke Fei Pang
- , Barbora Pardubska
- , Naomi Park
- , Aaron Parmar
- , Gaurang Patel
- , Minal Patel
- , Maggie Payne
- , Sharon Peacock
- , Arabella Petersen
- , Deborah Plowman
- , Tom Preston
- , Liam Prestwood
- , Christoph Puethe
- , Michael Quail
- , Diana Rajan
- , Shavanti Rajatileka
- , Richard Rance
- , Suzannah Rawlings
- , Nicholas Redshaw
- , Joe Reynolds
- , Mark Reynolds
- , Simon Rice
- , Matt Richardson
- , Connor Roberts
- , Katrina Robinson
- , Melanie Robinson
- , David Robinson
- , Hazel Rogers
- , Eduardo Martin Rojo
- , Daljit Roopra
- , Mark Rose
- , Luke Rudd
- , Ramin Sadri
- , Nicholas Salmon

- , David Saul
- , Frank Schwach
- , Carol Scott
- , Phil Seekings
- , Lesley Shirley
- , John Sillitoe
- , Alison Simms
- , Matthew Sinnott
- , Shanthi Sivadasan
- , Bart Siwek
- , Dale Sizer
- , Kenneth Skeldon
- , Jason Skelton
- , Joanna Slater-Tunstill
- , Lisa Sloper
- , Nathalie Smerdon
- , Chris Smith
- , Christen Smith
- , James Smith
- , Katie Smith
- , Michelle Smith
- , Sean Smith
- , Tina Smith
- , Leighton Sneade
- , Carmen Diaz Soria
- , Catarina Sousa
- , Emily Souster
- , Andrew Sparkes
- , Michael Spencer-Chapman
- , Janet Squares
- , Robert Stanley
- , Claire Steed
- , Tim Stickland
- , Ian Still
- , Michael R. Stratton
- , Michelle Strickland
- , Allen Swann
- , Agnieszka Swiatkowska
- , Neil Sycamore
- , Emma Swift
- , Edward Symons
- , Suzanne Szluha

- , Emma Taluy
- , Nunu Tao
- , Katy Taylor
- , Sam Taylor
- , Stacey Thompson
- , Mark Thompson
- , Mark Thomson
- , Nicholas Thomson
- , Scott Thurston
- , Gerry Tonkin-Hill
- , Dee Toombs
- , Benjamin Topping
- , Jaime Tovar-Corona
- , Daniel Ungureanu
- , James Uphill
- , Jana Urbanova
- , Philip Jansen Van Vuuren
- , Valerie Vancollie
- , Paul Voak
- , Danielle Walker
- , Matthew Walker
- , Matt Waller
- , Gary Ward
- , Charlie Weatherhogg
- , Niki Webb
- , Danni Weldon
- , Alan Wells
- , Eloise Wells
- , Luke Westwood
- , Theo Whipp
- , Thomas Whiteley
- , Georgia Whitton
- , Andrew Whitwham
- , Sara Widaa
- , Mia Williams
- , Mark Wilson
- & Sean Wright

**The COVID-19 Genomics UK (COG-UK) Consortium\***

- **Funding acquisition, leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, software and analysis tools, and visualization**
  - Samuel C. Robson
- **Funding acquisition, leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - Thomas R. Connor
  - & Nicholas J. Loman
- **Leadership and supervision, metadata curation, project administration, samples and logistics, sequencing and analysis, software and analysis tools, and visualization**
  - Tanya Golubchik
- **Funding acquisition, leadership and supervision, metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Rocio T. Martinez Nunez
- **Funding acquisition, leadership and supervision, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - David Bonsall
- **Funding acquisition, leadership and supervision, project administration, sequencing and analysis, software and analysis tools, and visualization**

- Andrew Rambaut
- **Funding acquisition, metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
- Luke B. Snell
- **Leadership and supervision, metadata curation, project administration, samples and logistics, software and analysis tools, and visualization**
- Rich Livett
- **Funding acquisition, leadership and supervision, metadata curation, project administration, and samples and logistics**
- Catherine Ludden
- **Funding acquisition, leadership and supervision, metadata curation, samples and logistics, and sequencing and analysis**
- Sally Corden
- & Eleni Nastouli
- **Funding acquisition, leadership and supervision, metadata curation, sequencing and analysis, and software and analysis tools**
- Gaia Nebbia
- **Funding acquisition, leadership and supervision, project administration, samples and logistics, and sequencing and analysis**
- Ian Johnston

- **Leadership and supervision, metadata curation, project administration, samples and logistics, and sequencing and analysis**
  - Katrina Lythgoe
  - , M. Estee Torok
  - & Ian G. Goodfellow
- **Leadership and supervision, metadata curation, project administration, samples and logistics, and visualization**
  - Jacqui A. Prieto
  - & Kordo Saeed
- **Leadership and supervision, metadata curation, project administration, sequencing and analysis, and software and analysis tools**
  - David K. Jackson
- **Leadership and supervision, metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Catherine Houlihan
- **Leadership and supervision, metadata curation, sequencing and analysis, software and analysis tools, and visualization**
  - Dan Frampton
- **Metadata curation, project administration, samples and logistics, sequencing and analysis, and software and analysis tools**
  - William L. Hamilton
  - & Adam A. Witney

- **Funding acquisition, samples and logistics, sequencing and analysis, and visualization**
  - Giselda Bucca
- **Funding acquisition, leadership and supervision, metadata curation and project administration**
  - Cassie F. Pope
- **Funding acquisition, leadership and supervision, metadata curation, and samples and logistics**
  - Catherine Moore
- **Funding acquisition, leadership and supervision, metadata curation, and sequencing and analysis**
  - Emma C. Thomson
- **Funding acquisition, leadership and supervision, project administration, and samples and logistics**
  - Ewan M. Harrison
- **Funding acquisition, leadership and supervision, sequencing and analysis, and visualization**
  - Colin P. Smith
- **Leadership and supervision, metadata curation, project administration, and sequencing and analysis**
  - Fiona Rogan

- **Leadership and supervision, metadata curation, project administration, and samples and logistics**
  - Shaun M. Beckwith
  - , Abigail Murray
  - , Dawn Singleton
  - , Kirstine Eastick
  - , Liz A. Sheridan
  - , Paul Randell
  - , Leigh M. Jackson
  - , Cristina V. Ariani
  - & Sónia Gonçalves
- **Leadership and supervision, metadata curation, samples and logistics, and sequencing and analysis**
  - Derek J. Fairley
  - , Matthew W. Loose
  - & Joanne Watkins
- **Leadership and supervision, metadata curation, samples and logistics, and visualization**
  - Samuel Moses
- **Leadership and supervision, metadata curation, sequencing and analysis, and software and analysis tools**
  - Sam Nicholls
  - , Matthew Bull
  - & Roberto Amato
- **Leadership and supervision, project administration, samples and logistics, and sequencing and analysis**
  - Darren L. Smith

- **Leadership and supervision, sequencing and analysis, software and analysis tools, and visualization**
  - David M. Aanensen
  - & Jeffrey C. Barrett
- **Metadata curation, project administration, samples and logistics, and sequencing and analysis**
  - Dinesh Aggarwal
  - , James G. Shepherd
  - , Martin D. Curran
  - & Surendra Parmar
- **Metadata curation, project administration, sequencing and analysis, and software and analysis tools**
  - Matthew D. Parker
- **Metadata curation, samples and logistics, sequencing and analysis, and software and analysis tools**
  - Catryn Williams
- **Metadata curation, samples and logistics, sequencing and analysis, and visualization**
  - Sharon Glaysher
- **Metadata curation, sequencing and analysis, software and analysis tools, and visualization**
  - Anthony P. Underwood
  - , Matthew Bashton
  - , Nicole Pacchiarini

- , Katie F. Loveson
- & Matthew Byott
- **Project administration, sequencing and analysis, software and analysis tools, and visualization**
  - Alessandro M. Carabelli
- **Funding acquisition, leadership and supervision, and metadata curation**
  - Kate E. Templeton

- **Funding acquisition, leadership and supervision, and project administration**
  - Thushan I. de Silva
  - , Dennis Wang
  - , Cordelia F. Langford
  - & John Sillitoe
- **Funding acquisition, leadership and supervision, and samples and logistics**
  - Rory N. Gunson
- **Funding acquisition, leadership and supervision, and sequencing and analysis**
  - Simon Cottrell
  - , Justin O'Grady
  - & Dominic Kwiatkowski
- **Leadership and supervision, metadata curation and project administration**
  - Patrick J. Lillie
- **Leadership and supervision, metadata curation, and samples and logistics**
  - Nicholas Cortes
  - , Nathan Moore
  - , Claire Thomas
  - , Phillipa J. Burns
  - , Tabitha W. Mahungu
  - & Steven Liggett

- **Leadership and supervision, metadata curation, and sequencing and analysis**
  - Angela H. Beckett
  - & Matthew T. G. Holden
- **Leadership and supervision, project administration, and samples and logistics**
  - Lisa J. Levett
  - , Husam Osman
  - & Mohammed O. Hassan-Ibrahim
- **Leadership and supervision, project administration, and sequencing and analysis**
  - David A. Simpson
- **Leadership and supervision, samples and logistics, and sequencing and analysis**
  - Meera Chand
  - , Ravi K. Gupta
  - , Alistair C. Darby
  - & Steve Paterson
- **Leadership and supervision, sequencing and analysis, and software and analysis tools**
  - Oliver G. Pybus
  - , Erik M. Volz
  - , Daniela de Angelis
  - , David L. Robertson
  - , Andrew J. Page
  - & Inigo Martincorena

- **Leadership and supervision, sequencing and analysis, and visualization**
  - Louise Aigrain
  - & Andrew R. Bassett
- **Metadata curation, project administration, and samples and logistics**
  - Nick Wong
  - , Yusri Taha
  - , Michelle J. Erkert
  - & Michael H. Spencer Chapman
- **Metadata curation, project administration, and sequencing and analysis**
  - Rebecca Dewar
  - & Martin P. McHugh
- **Metadata curation, project administration, and software and analysis tools**
  - Siddharth Mookerjee
- **Metadata curation, project administration and visualization**
  - Stephen Aplin
  - , Matthew Harvey
  - , Thea Sass
  - , Helen Umpleby
  - & Helen Wheeler
- **Metadata curation, samples and logistics, and sequencing and analysis**

- James P. McKenna
- , Ben Warne
- , Joshua F. Taylor
- , Yasmin Chaudhry
- , Rhys Izuagbe
- , Aminu S. Jahun
- , Gregory R. Young
- , Claire McMurray
- , Clare M. McCann
- , Andrew Nelson
- & Scott Elliott

- **Metadata curation, samples and logistics, and visualization**
  - Hannah Lowe
- **Metadata curation, sequencing and analysis, and software and analysis tools**
  - Anna Price
  - , Matthew R. Crown
  - , Sara Rey
  - , Sunando Roy
  - & Ben Temperton
- **Metadata curation, sequencing and analysis, and visualization**
  - Sharif Shaaban
  - & Andrew R. Hesketh
- **Project administration, samples and logistics, and sequencing and analysis**
  - Kenneth G. Laing
  - , Irene M. Monahan
  - & Judith Heaney

- **Project administration, samples and logistics, and visualization**
  - Emanuela Pelosi
  - , Siona Silviera
  - & Eleri Wilson-Davies
- **Samples and logistics, software and analysis tools, and visualization**
  - Helen Fryer
- **Sequencing and analysis, software and analysis tools, and visualization**
  - Helen Adams
  - , Louis du Plessis
  - , Rob Johnson
  - , William T. Harvey
  - , Joseph Hughes
  - , Richard J. Orton
  - , Lewis G. Spurgin
  - , Yann Bourgeois
  - , Chris Ruis
  - , Áine O’Toole
  - , Marina Gourtovaia
  - & Theo Sanderson
- **Funding acquisition, and leadership and supervision**
  - Christophe Fraser
  - , Jonathan Edgeworth
  - , Judith Breuer
  - , Stephen L. Michell
  - & John A. Todd
- **Funding acquisition and project administration**

- Michaela John
- & David Buck
- **Leadership and supervision, and metadata curation**
  - Kavitha Gajee
  - & Gemma L. Kay
- **Leadership and supervision, and project administration**
  - Sharon J. Peacock
  - & David Heyburn
- **Leadership and supervision, and samples and logistics**
  - Katie Kitchman
  - , Alan McNally
  - , David T. Pritchard
  - , Samir Dervisevic
  - , Peter Muir
  - , Esther Robinson
  - , Barry B. Vipond
  - , Newara A. Ramadan
  - , Christopher Jeanes
  - , Danni Weldon
  - , Jana Catalan
  - & Neil Jones
- **Leadership and supervision, and sequencing and analysis**
  - Ana da Silva Filipe
  - , Chris Williams
  - , Marc Fuchs
  - , Julia Miskelly
  - , Aaron R. Jeffries
  - , Karen Oliver

- & Naomi R. Park
- **Metadata curation, and samples and logistics**
  - Amy Ash
  - , Cherian Koshy
  - , Magdalena Barrow
  - , Sarah L. Buchan
  - , Anna Mantzouratou
  - , Gemma Clark
  - , Christopher W. Holmes
  - , Sharon Campbell
  - , Thomas Davis
  - , Ngee Keong Tan
  - , Julianne R. Brown
  - , Kathryn A. Harris
  - , Stephen P. Kidd
  - , Paul R. Grant
  - , Li Xu-McCrae
  - , Alison Cox
  - , Pinglawathee Madona
  - , Marcus Pond
  - , Paul A. Randell
  - , Karen T. Withell
  - , Cheryl Williams
  - , Clive Graham
  - , Rebecca Denton-Smith
  - , Emma Swindells
  - , Robyn Turnbull
  - , Tim J. Sloan
  - , Andrew Bosworth
  - , Stephanie Hutchings
  - , Hannah M. Pymont
  - , Anna Casey
  - , Liz Ratcliffe
  - , Christopher R. Jones
  - , Bridget A. Knight

- , Tanzina Haque
- , Jennifer Hart
- , Dianne Irish-Tavares
- , Eric Witele
- , Craig Mower
- , Louisa K. Watson
- , Jennifer Collins
- , Gary Eltringham
- , Dorian Crudgington
- , Ben Macklin
- , Miren Iturriza-Gomara
- , Anita O. Lucaci
- & Patrick C. McClure

- **Metadata curation, and sequencing and analysis**

- Matthew Carlile
- , Nadine Holmes
- , Christopher Moore
- , Nathaniel Storey
- , Stefan Rooke
- , Gonzalo Yebra
- , Noel Craine
- , Malorie Perry
- , Nabil-Fareed Alikhan
- , Stephen Bridgett
- , Kate F. Cook
- , Christopher Fearn
- , Salman Goudarzi
- , Ronan A. Lyons
- , Thomas Williams
- , Sam T. Haldenby
- , Jillian Durham
- & Steven Leonard

- **Metadata curation, and software and analysis tools**

- Robert M. Davies

- **Project administration, and samples and logistics**

- Rahul Batra
- , Beth Blane
- , Moira J. Spyer
- , Perminder Smith
- , Mehmet Yavus
- , Rachel J. Williams
- , Adhyana I. K. Mahanama
- , Buddhini Samaraweera
- , Sophia T. Girgis
- , Samantha E. Hansford
- , Angie Green
- , Charlotte Beaver
- , Katherine L. Bellis
- , Matthew J. Dorman
- , Sally Kay
- , Liam Prestwood
- & Shavanti Rajatileka

- **Project administration, and sequencing and analysis**

- Joshua Quick

- **Project administration, and software and analysis tools**

- Radoslaw Poplawski

- **Samples and logistics, and sequencing and analysis**

- Nicola Reynolds
- , Andrew Mack
- , Arthur Morriss
- , Thomas Whalley

- , Bindi Patel
- , Iliana Georgana
- , Myra Hosmillo
- , Malte L. Pinckert
- , Joanne Stockton
- , John H. Henderson
- , Amy Hollis
- , William Stanley
- , Wen C. Yew
- , Richard Myers
- , Alicia Thornton
- , Alexander Adams
- , Tara Annett
- , Hibo Asad
- , Alec Birchley
- , Jason Coombes
- , Johnathan M. Evans
- , Laia Fina
- , Bree Gatica-Wilcox
- , Lauren Gilbert
- , Lee Graham
- , Jessica Hey
- , Ember Hilvers
- , Sophie Jones
- , Hannah Jones
- , Sara Kumziene-Summerhayes
- , Caoimhe McKerr
- , Jessica Powell
- , Georgia Pugh
- , Sarah Taylor
- , Alexander J. Trotter
- , Charlotte A. Williams
- , Leanne M. Kermack
- , Benjamin H. Foulkes
- , Marta Gallis
- , Hailey R. Hornsby
- , Stavroula F. Louka

- , Manoj Pohare
- , Paige Wolverson
- , Peijun Zhang
- , George MacIntyre-Cockett
- , Amy Trebes
- , Robin J. Moll
- , Lynne Ferguson
- , Emily J. Goldstein
- , Alasdair Maclean
- & Rachael Tomb

- **Samples and logistics, and software and analysis tools**

- Igor Starinskij

- **Sequencing and analysis, and software and analysis tools**

- Laura Thomson
- , Joel Southgate
- , Moritz U. G. Kraemer
- , Jayna Raghwani
- , Alex E. Zarebski
- , Olivia Boyd
- , Lily Geidelberg
- , Chris J. Illingworth
- , Chris Jackson
- , David Pascall
- , Sreenu Vattipally
- , Timothy M. Freeman
- , Sharon N. Hsu
- , Benjamin B. Lindsey
- , Keith James
- , Kevin Lewis
- , Gerry Tonkin-Hill
- & Jaime M. Tovar-Corona

- **Sequencing and analysis, and visualization**
  - MacGregor Cox
- **Software and analysis tools, and visualization**
  - Khalil Abudahab
  - , Mirko Menegazzo
  - , Ben E. W. Taylor
  - , Corin A. Yeats
  - , Afrida Mukaddas
  - , Derek W. Wright
  - , Leonardo de Oliveira Martins
  - , Rachel Colquhoun
  - , Verity Hill
  - , Ben Jackson
  - , J. T. McCrone
  - , Nathan Medd
  - , Emily Scher
  - & Jon-Paul Keatley
- **Leadership and supervision**
  - Tanya Curran
  - , Sian Morgan
  - , Patrick Maxwell
  - , Ken Smith
  - , Sahar Eldirdiri
  - , Anita Kenyon
  - , Alison H. Holmes
  - , James R. Price
  - , Tim Wyatt
  - , Alison E. Mather
  - , Timofey Skvortsov
  - & John A. Hartley

- **Metadata curation**

- Martyn Guest
- , Christine Kitchen
- , Ian Merrick
- , Robert Munn
- , Beatrice Bertolusso
- , Jessica Lynch
- , Gabrielle Vernet
- , Stuart Kirk
- , Elizabeth Wastnedge
- , Rachael Stanley
- , Giles Idle
- , Declan T. Bradley
- , Jennifer Poyner
- & Matilde Mori

- **Project administration**

- Owen Jones
- , Victoria Wright
- , Ellena Brooks
- , Carol M. Churcher
- , Mireille Fragakis
- , Katerina Galai
- , Andrew Jermy
- , Sarah Judges
- , Georgina M. McManus
- , Kim S. Smith
- , Elaine Westwick
- , Stephen W. Attwood
- , Frances Bolt
- , Alisha Davies
- , Elen De Lacy
- , Fatima Downing
- , Sue Edwards
- , Lizzie Meadows

- , Sarah Jeremiah
- , Nikki Smith
- & Luke Foulser

- **Samples and logistics**

- Themoula Charalampous
- , Amita Patel
- , Louise Berry
- , Tim Boswell
- , Vicki M. Fleming
- , Hannah C. Howson-Wells
- , Amelia Joseph
- , Manjinder Khakh
- , Michelle M. Lister
- , Paul W. Bird
- , Karlie Fallon
- , Thomas Helmer
- , Claire L. McMurray
- , Mina Odedra
- , Jessica Shaw
- , Julian W. Tang
- , Nicholas J. Willford
- , Victoria Blakey
- , Veena Raviprakash
- , Nicola Sheriff
- , Lesley-Anne Williams
- , Theresa Feltwell
- , Luke Bedford
- , James S. Cargill
- , Warwick Hughes
- , Jonathan Moore
- , Susanne Stonehouse
- , Laura Atkinson
- , Jack C. D. Lee
- , Divya Shah
- , Adela Alcolea

- , Natasha Ohemeng-Kumi
- , John Ramble
- , Jasveen Sehmi
- , Rebecca Williams
- , Wendy Chatterton
- , Monika Pusok
- , William Everson
- , Anibolina Castigador
- , Emily Macnaughton
- , Kate El Bouzidi
- , Temi Lampejo
- , Malur Sudhanva
- , Cassie Breen
- , Graciela Sluga
- , Shazaad S. Y. Ahmad
- , Ryan P. George
- , Nicholas W. Machin
- , Debbie Binns
- , Victoria James
- , Rachel Blacow
- , Lindsay Coupland
- , Louise Smith
- , Edward Barton
- , Debra Padgett
- , Garren Scott
- , Aidan Cross
- , Mariyam Mirfenderesky
- , Jane Greenaway
- , Kevin Cole
- , Phillip Clarke
- , Nichola Duckworth
- , Sarah Walsh
- , Kelly Bicknell
- , Robert Impey
- , Sarah Wyllie
- , Richard Hopes
- , Chloe Bishop

- , Vicki Chalker
- , Ian Harrison
- , Laura Gifford
- , Zoltan Molnar
- , Cressida Auckland
- , Cariad Evans
- , Kate Johnson
- , David G. Partridge
- , Mohammad Raza
- , Paul Baker
- , Stephen Bonner
- , Sarah Essex
- , Leanne J. Murray
- , Andrew I. Lawton
- , Shirelle Burton-Fanning
- , Brendan A. I. Payne
- , Sheila Waugh
- , Andrea N. Gomes
- , Maimuna Kimuli
- , Darren R. Murray
- , Paula Ashfield
- , Donald Dobie
- , Fiona Ashford
- , Angus Best
- , Liam Crawford
- , Nicola Cumley
- , Megan Mayhew
- , Oliver Megram
- , Jeremy Mirza
- , Emma Moles-Garcia
- , Benita Percival
- , Leah Ensell
- , Helen L. Lowe
- , Laurentiu Maftei
- , Matteo Mondani
- , Nicola J. Chaloner
- , Benjamin J. Cogger

- , Lisa J. Easton
- , Hannah Huckson
- , Jonathan Lewis
- , Sarah Lowdon
- , Cassandra S. Malone
- , Florence Munemo
- , Manasa Mutingwende
- , Roberto Nicodemi
- , Olga Podplomyk
- , Thomas Somassa
- , Andrew Beggs
- , Alex Richter
- , Claire Cormie
- , Joana Dias
- , Sally Forrest
- , Ellen E. Higginson
- , Mailis Maes
- , Jamie Young
- , Rose K. Davidson
- , Kathryn A. Jackson
- , Lance Turtle
- , Alexander J. Keeley
- , Jonathan Ball
- , Timothy Byaruhanga
- , Joseph G. Chappell
- , Jayasree Dey
- , Jack D. Hill
- , Emily J. Park
- , Arezou Fanaie
- , Rachel A. Hilson
- , Geraldine Yaze
- & Stephanie Lo

- **Sequencing and analysis**

- Safiah Afifi
- , Robert Beer

- , Joshua Maksimovic
- , Kathryn McCluggage Masters
- , Karla Spellman
- , Catherine Bresner
- , William Fuller
- , Angela Marchbank
- , Trudy Workman
- , Ekaterina Shelest
- , Johnny Debebe
- , Fei Sang
- , Marina Escalera Zamudio
- , Sarah Francois
- , Bernardo Gutierrez
- , Tetyana I. Vasylyeva
- , Flavia Flaviani
- , Manon Ragonnet-Cronin
- , Katherine L. Smollett
- , Alice Broos
- , Daniel Mair
- , Jenna Nichols
- , Kyriaki Nomikou
- , Lily Tong
- , Ioulia Tsatsani
- , Sarah O'Brien
- , Steven Rushton
- , Roy Sanderson
- , Jon Perkins
- , Seb Cotton
- , Abbie Gallagher
- , Elias Allara
- , Clare Pearson
- , David Bibby
- , Gavin Dabrera
- , Nicholas Ellaby
- , Eileen Gallagher
- , Jonathan Hubb
- , Angie Lackenby

- , David Lee
- , Nikos Manesis
- , Tamyo Mbisa
- , Steven Platt
- , Katherine A. Twohig
- , Mari Morgan
- , Alp Aydin
- , David J. Baker
- , Ebenezer Foster-Nyarko
- , Sophie J. Prosolek
- , Steven Rudder
- , Chris Baxter
- , Sílvia F. Carvalho
- , Deborah Lavin
- , Arun Mariappan
- , Clara Radulescu
- , Aditi Singh
- , Miao Tang
- , Helen Morcrette
- , Nadua Bayzid
- , Marius Cotic
- , Carlos E. Balcazar
- , Michael D. Gallagher
- , Daniel Maloney
- , Thomas D. Stanton
- , Kathleen A. Williamson
- , Robin Manley
- , Michelle L. Michelsen
- , Christine M. Sambles
- , David J. Studholme
- , Joanna Warwick-Dugdale
- , Richard Eccles
- , Matthew Gemmell
- , Richard Gregory
- , Margaret Hughes
- , Charlotte Nelson
- , Lucille Rainbow

- , Edith E. Vamos
- , Hermione J. Webster
- , Mark Whitehead
- , Claudia Wierzbicki
- , Adrienn Angyal
- , Luke R. Green
- , Max Whiteley
- , Emma Betteridge
- , Iraad F. Bronner
- , Ben W. Farr
- , Scott Goodwin
- , Stefanie V. Lensing
- , Shane A. McCarthy
- , Michael A. Quail
- , Diana Rajan
- , Nicholas M. Redshaw
- , Carol Scott
- , Lesley Shirley
- & Scott A. J. Thurston

- **Software and analysis tools**

- Will Rowe
- , Amy Gaskin
- , Thanh Le-Viet
- , James Bonfield
- , Jennifer Liddle
- & Andrew Whitwham

## **Contributions**

H.S.V. and M.G. developed the analysis code, which H.S.V. implemented with input from A.W.J.; H.S.V. created most of the figures. M.S. analysed, annotated and aggregated viral genome data. N.D.M. conducted phylogeographic analyses supervised by N.G.; T.S., R.G., M.S. and H.S.V. developed the interactive spatiotemporal viewer. T.N., F.S., I.H., R.A., C.A., S.G., D.J., I.J., C.S., J.S., T.S. and M.S. analysed genomic surveillance data

under the supervision of D.K., M.C., I.M. and J.C.B.; J.H. and S.F. analysed ONS data and helped with epidemiological modelling and data interpretation. E.V. analysed growth rates and helped with data interpretation. E.B. and J.P.G. supervised H.S.V. and helped with data interpretation. J.C.B. and M.G. supervised the analysis with advice from I.M.; M.G., H.S.V., M.S., N.D.M., T.S., I.M. and J.C.B. wrote the manuscript with input from all of the co-authors.

## Corresponding authors

Correspondence to [Jeffrey C. Barrett](#) or [Moritz Gerstung](#).

## Ethics declarations

## Competing interests

E.B. is a paid consultant of Oxford Nanopore.

## Additional information

**Peer review information** *Nature* thanks Tulio De Oliveira, Philippe Lemey and Matthew Scotch for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 SARS-CoV-2 surveillance sequencing in England between September 2020 and June 2021.](#)

- a. Local monthly coverage across 315 LTLAs. b. Weekly coverage of genomic surveillance sequencing. c. Hospitalization, case and infection

fatality rates relative to ONS prevalence. Dots denote mean estimates and error bars 95% CIs.

### Extended Data Fig. 2 Genomic surveillance model of total incidence and lineage-specific frequencies.

**a.** Cubic basis splines (top row) are convolved with the infection to test distribution (row 2 and 3) and used to fit the log incidence in a LTLA and its corresponding derivatives (growth rates; bottom row). **b.** Example incidence (top row), logarithmic incidence with individual convolved basis functions (dashed lines, row 2), growth rate with individual spline basis derivatives (dashed lines, row 3) and resulting (case) reproduction numbers (growth rate per 5.1d) from our approach (GenomicSurveillance) and estimates by EpiEstim<sup>48</sup>, shifted by 10d to approximate a case reproduction number. **c.** The relative frequencies of 62 different lineages are modelled using piecewise multinomial logistic regression. The linear logits are modelled to jump stochastically within 21d prior to first observation to account for the effects of new introductions. Shown are the logits of 5 selected lineages in two different LTLAs.

### Extended Data Fig. 3 Genomic surveillance model selection.

**a.** Model loss in terms of the ELBO objective function and the model hyperparameters alpha0 and alpha1 (see [Methods](#)). **b.** Model deviance (calculated as  $-2 \times \log$  pointwise predictive density) with respect to the model hyperparameters  $\alpha_0$  and  $\alpha_1$  (see [Methods](#)). **c.** Mean squared error (MSE) of modelled weekly proportions of highly prevalent lineages with respect to the model parameters  $\alpha_0$  and  $\alpha_1$  (see [Methods](#)). **d.** Same as in **c**, but for lineages exhibiting low frequencies (VOCs).

### Extended Data Fig. 4 Spatiotemporal model of 71 SARS-CoV-2 lineages in 315 English LTLAs between September 2020 and June 2021.

**a.** Regional lineage specific relative frequency of lineages contributing more than 50 genomes during the time period shown. Dots denote observed

data, lines the fits aggregated to each region. **b.** Same as **a**, but on a log scale. **c.** Same data as in **a**, shown as stacked bar charts. Colours resemble major lineages as indicated and shadings thereof indicate sublineages. **d.** Same fits as in **a**, shown as stacked segments. **e.** Average growth rates for 71 SARS-CoV2 lineages estimated in different regions in England. Dots denote median estimates and error bars 95% CIs.

### Extended Data Fig. 5 Relative growth of B.1.177.

**a.** Lineage-specific relative frequency data in England, excluding B.1.1.7 and other VOCs/VUIs (Category Other includes: A, A.18, A.20, A.23, A.25, A.27, A.28, B, B.29, B.40, None). Colours resemble major lineages as indicated and shadings thereof indicate sublineages. **b.** Lineage-specific relative frequency data in Denmark, excluding B.1.1.7 and other VOCs/VUIs. Colours resemble major lineages as indicated and shadings thereof indicate sublineages.

### Extended Data Fig. 6 Genomic diversity of the SARS-CoV-2 epidemic.

Shown is the entropy (blue), total number of observed Pango lineages (grey, divided by 4), as well as the proportion of B.1.1.7 (orange, right axis). The sweep of B.1.1.7 causes an intermittent decline of genomic diversity as measured by the entropy.

### Extended Data Fig. 7 Global phylogenetic trees of selected VOCs/VUIs.

English surveillance and other (targeted and quarantine) samples are highlighted respectively orange and red.

### Extended Data Fig. 8 Global phylogenetic trees of B.1.617 sublineages.

**a, b and c.** English surveillance and other (targeted and quarantine) samples are highlighted respectively orange and red. The trees of B.1.617.1 and

B.1.617.2 are rooted. **d**. Number of UK introductions inferred by parsimony (minimum and maximum numbers) and by Thorne BEAST (95% posterior CI) for each VOC.

## Supplementary information

### Supplementary Notes

Supplementary Notes 1–3.

### Reporting Summary

### Peer Review File

### Supplementary Tables

Supplementary Tables 1–4.

## Source data

### Source Data Fig. 1

### Source Data Fig. 2

### Source Data Fig. 3

### Source Data Fig. 4

### Source Data Fig. 5

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## Reprints and Permissions

## About this article

### Cite this article

Vöhringer, H.S., Sanderson, T., Sinnott, M. *et al.* Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature* **600**, 506–511 (2021).  
<https://doi.org/10.1038/s41586-021-04069-y>

- Received: 22 May 2021
- Accepted: 29 September 2021
- Published: 14 October 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04069-y>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [Implications of testicular ACE2 and the renin–angiotensin system for SARS-CoV-2 on testis function](#)
  - R. Clayton Edenfield
  - Charles A. Easley

*Nature Reviews Urology* (2021)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04069-y>.

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 20 September 2021](#)

# High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape

- [Fabian Schmidt](#)<sup>1,nal</sup>,
- [Yiska Weisblum](#)<sup>1,nal</sup>,
- [Magdalena Rutkowska](#)<sup>2</sup>,
- [Daniel Poston](#) [ORCID: orcid.org/0000-0001-9900-6717](#)<sup>1</sup>,
- [Justin DaSilva](#) [ORCID: orcid.org/0000-0001-6393-0000](#)<sup>1</sup>,
- [Fengwen Zhang](#)<sup>1</sup>,
- [Eva Bednarski](#)<sup>1</sup>,
- [Alice Cho](#)<sup>3</sup>,
- [Dennis J. Schaefer-Babajew](#) [ORCID: orcid.org/0000-0002-5380-2950](#)<sup>3</sup>,
- [Christian Gaebler](#) [ORCID: orcid.org/0000-0001-7295-8128](#)<sup>3</sup>,
- [Marina Caskey](#) [ORCID: orcid.org/0000-0003-1727-8693](#)<sup>3</sup>,
- [Michel C. Nussenzweig](#) [ORCID: orcid.org/0000-0003-0592-8564](#)<sup>2,3</sup>,
- [Theodora Hatzioannou](#) [ORCID: orcid.org/0000-0002-7889-0766](#)<sup>1</sup> &
- [Paul D. Bieniasz](#) [ORCID: orcid.org/0000-0002-2368-3719](#)<sup>1,2</sup>

*Nature* volume 600, pages 512–516 (2021)

- 104k Accesses
- 5 Citations
- 2294 Altmetric
- [Metrics details](#)

## Subjects

- [Immune evasion](#)
- [SARS-CoV-2](#)

## Abstract

The number and variability of the neutralizing epitopes targeted by polyclonal antibodies in individuals who are SARS-CoV-2 convalescent and vaccinated are key determinants of neutralization breadth and the genetic barrier to viral escape<sup>1,2,3,4</sup>. Using HIV-1 pseudotypes and plasma selection experiments with vesicular stomatitis virus/SARS-CoV-2 chimaeras<sup>5</sup>, here we show that multiple neutralizing epitopes, within and outside the receptor-binding domain, are variably targeted by human polyclonal antibodies. Antibody targets coincide with spike sequences that are enriched for diversity in natural SARS-CoV-2 populations. By combining plasma-selected spike substitutions, we generated synthetic ‘polymutant’ spike protein pseudotypes that resisted polyclonal antibody neutralization to a similar degree as circulating variants of concern. By aggregating variant of concern-associated and antibody-selected spike substitutions into a single polymutant spike protein, we show that 20 naturally occurring mutations in the SARS-CoV-2 spike protein are sufficient to generate pseudotypes with near-complete resistance to the polyclonal neutralizing antibodies generated by individuals who are convalescent or recipients who received an mRNA vaccine. However, plasma from individuals who had been infected and subsequently received mRNA vaccination neutralized pseudotypes bearing this highly resistant SARS-CoV-2 polymutant spike, or diverse sarbecovirus spike proteins. Thus, optimally elicited human polyclonal antibodies against SARS-CoV-2 should be resilient to substantial future SARS-CoV-2 variation and may confer protection against potential future sarbecovirus pandemics.

[Download PDF](#)

## Main

Neutralizing antibodies elicited by previous infection or by vaccination probably represent a key component of protective immunity against SARS-CoV-2. Antibodies targeting the receptor-binding domain (RBD) of the spike protein are thought to dominate the neutralizing activity of convalescent or vaccine recipient plasma<sup>6</sup>, and include the most potent neutralizing antibodies cloned from individuals who are convalescent<sup>7,8,9,10,11</sup>. Nevertheless, additional SARS-CoV-2 neutralizing antibody targets include the N-terminal domain (NTD) and the fusion machinery<sup>4,8,10,12,13</sup>, and the full spectrum of epitopes targeted by neutralizing antibodies in convalescent or vaccine recipient plasma has not been defined. SARS-CoV-2 variants of concern (VOCs) or variants of interest (VOIs) encode spike amino acid substitutions<sup>14,15,16,17</sup>, some of which confer resistance to individual human monoclonal antibodies but have variable, typically modest, effects on neutralization by polyclonal plasma antibodies<sup>1,6,9,17,18,19,20</sup>. Mutated sites in VOCs include those in the RBD, NTD and elsewhere, but the numbers and locations of spike substitutions required for SARS-CoV-2 to evade the polyclonal neutralizing antibodies encountered in recipients of a

vaccine or who are convalescent are unknown and are crucial determinants of population immunity.

## Polyclonal neutralizing antibody targets

Exploiting the fact that SARS-CoV is poorly neutralized by SARS-CoV-2 convalescent plasma, we compared the sensitivity of HIV-1 pseudotypes bearing parental and chimeric spike proteins in which RBD sequences were exchanged (SARS-CoV-2(1-RBD) and SARS-CoV(2-RBD); Fig. 1a, Extended Data Fig. 1) to neutralization by plasma from 26 individuals from a previously described Rockefeller University COVID-19 convalescent cohort<sup>21</sup>. The plasma samples were obtained at an average of 1.3 months after infection and were selected for high SARS-CoV-2 neutralization titres (the RU27 plasma panel). Compared with the SARS-CoV-2 pseudotype, the SARS-CoV-2(1-RBD) pseudotype was less sensitive to neutralization by 21 of the 26 plasmas (median difference = 1.8-fold, range 0.5–9.8-fold,  $P = 0.0005$  (Wilcoxon two-tailed test); Fig. 1b). Conversely, the SARS-CoV(2-RBD) pseudotype was more sensitive than the SARS-CoV pseudotype to all plasmas (median difference = 8-fold, range 1.2–75.5-fold,  $P < 0.0001$  (Wilcoxon two-tailed test); Fig 1c). Nevertheless, the neutralizing potency of some plasmas was hardly affected when the SARS-CoV-2 RBD was replaced by the SARS-CoV RBD, even though some of those plasmas were minimally active against SARS-CoV (for example, RU9, RU10, RU11 and RU15; Fig. 1b,c). Indeed, plasmas that poorly neutralized SARS-CoV potently neutralized chimeric spike pseudotypes with either RBD or the other spike domains from SARS-CoV-2 (Fig. 1b,c). For the plasma panel as a whole, pseudotype neutralizing potency against SARS-CoV-2 did not correlate with potency against SARS-CoV or SARS-CoV-2(1-RBD), but did correlate with potency against SARS-CoV(2-RBD) (Extended Data Fig. 2a–c). Although altered conformation of the RBD in chimeric spike proteins might affect neutralization, these data indicate that while the RBD constitutes a major neutralizing target, substantial plasma neutralizing activity is also directed against non-RBD epitopes.

**Fig. 1: Neutralizing antibodies in SARS-CoV-2 convalescent plasma targeting both RBD and non-RBD determinants.**

---

 **figure1**

**a**, Design of RBD-exchanged chimeric spike proteins. **b, c**, NT<sub>50</sub> for 26 high-titre convalescent plasmas (from the RU1–27 panel) against pseudotyped HIV-1 virions bearing the indicated spike proteins. A median of two independent experiments is plotted. FP, fusion peptide; HR, heptad repeat; TM, transmembrane.

[Source data](#)

## Polyclonal antibody-selected variants

To more precisely map the targets of polyclonal neutralizing antibodies in individuals who are convalescent, we passaged a recombinant vesicular stomatitis virus (rVSV)/SARS-CoV-2 chimeric virus<sup>1,5</sup> in the presence of each of the RU27 plasmas for up to six passages. rVSV/SARS-CoV-2 mimics the neutralization properties of SARS-CoV-2 (refs. <sup>1,5</sup>) but obviates the safety concerns that would accompany such studies with authentic SARS-CoV-2. Next-generation sequencing indicated that rVSV/SARS-CoV-2 passage in 22 out of the 27 plasmas led to the selection of spike mutants (Fig. [2a](#), Extended Data Fig. [3](#), Supplementary Table 1). For some plasmas, multiple substitutions were selected at distinct but proximal sites in viral subpopulations, indicating a dominant neutralizing activity targeting a particular epitope. For other plasmas, substitutions were enriched in multiple regions of the spike coding sequence, suggesting co-dominant neutralizing activities. Six passages of rVSV/SARS-CoV-2 without plasma enriched a small number of substitutions that were assumed to represent cell-culture or fitness-enhancing adaptations (for example, T604I), but were distinct from the majority of substitutions arising after rVSV/SARS-CoV-2 passage in plasma (Extended Data Fig. [3](#), Supplementary Table 1). Cumulatively, the plasma-selected mutations were enriched in specific elements within the NTD, RBD and other spike domains (Fig. [2a](#), Supplementary Table 1). From the 27 plasma-passaged virus populations, 38 individual mutant viruses were isolated by plaque purification; each encoded one, two or three spike substitutions (Fig. [2b](#)) that

generally occurred at high frequency in the passaged viral populations (Supplementary Table 1).

**Fig. 2: Selection of SARS-CoV-2 spike mutants by polyclonal antibodies.**

 figure2



**a**, Frequencies of amino acid substitutions at each codon of the SARS-CoV-2 spike protein in two independent rVSV/SARS-CoV-2 populations (1D7 and 2E1), determined by Illumina sequencing. Pooled results following selection with the RU27 plasma panel are displayed. **b**, Locations of amino acid substitutions in 38 plaque-purified rVSV/SARS-CoV-2 isolates obtained from rVSV/SARS-CoV-2 populations following passage in the RU27 plasmas. **c**, Frequencies of naturally occurring amino acid substitutions (red circles) at each codon of the SARS-CoV-2 spike protein. The shaded grey bars in **a–c** indicate shared regions where variation is enriched. **d**, Comparison of the averaged frequency of substitutions observed after passaging rVSV/SARS-CoV-2 with RU27 plasmas (centre) and the frequency of sequence changes in natural populations (right), projected onto the SARS-CoV-2 spike structure (PDB 6VXX) with positions of the RBD and NTD indicated (left). The average frequency of substitutions in a 15 Å radius is represented using the colour spectrum (scale = 0–20 (centre) and 0–9 (right)). **e**, Neutralization potency of RU27 plasmas

against rVSV/SARS-CoV-2 encoding wild type (WT), individually selected mutants or PMS1-1 spike proteins. A median of two independent determinations is plotted.

### Source data

We compared the distribution of mutations selected by rVSV/SARS-CoV-2 passage with the RU27 plasma panel with those occurring in circulating SARS-CoV-2 populations (Fig. 2a–d). In both plasma-selected and naturally occurring sequences, substitutions were enriched in several elements that contribute to the ‘supersite’ targeted by NTD-binding neutralizing antibodies<sup>12,13</sup> (Fig. 2a–d). Similar plasma-selected and natural sequence variation was also evident in elements targeted by class 2 and class 3 RBD-binding neutralizing antibodies<sup>22</sup>. Mutations known to confer resistance to class I RBD antibodies were not selected by plasma passage, perhaps reflecting a lower than expected abundance of class I antibodies in this plasma panel (Fig. 2a–d). Other sites, including spike amino acids approximately 680–700 and approximately 930, exhibited variation in both plasma-passaged and natural variant datasets, but have not yet been demonstrated to be targeted by neutralizing antibodies. Nevertheless, the similarity in the distribution of natural and plasma-selected sequence variation within the spike protein suggests that selection by neutralizing antibodies is a driver of divergence in circulating SARS-CoV-2 populations.

Of the 38 plaque-purified rVSV/SARS-CoV-2 mutants recovered following passage in RU27 plasmas, 34 exhibited varying degrees of reduced sensitivity to neutralization by the plasma that was used for its selection (median 3.1-fold reduced half-maximal neutralizing titre (NT<sub>50</sub>), range 0.8–39.3-fold; Extended Data Figs. 4, 5). Nevertheless, for 37 of the 38 selected rVSV/SARS-CoV-2 mutants, the selecting plasma exhibited residual neutralizing activity. We aggregated 13 mutations from the plasma-selected viruses based on their effects on plasma neutralization sensitivity (Extended Data Figs. 4, 5) and distribution throughout the spike protein, generating a single synthetic ‘polymutant’ spike (PMS) protein sequence, termed PMS1-1 (Extended Data Fig. 6a). An rVSV/SARS-CoV-2 derivative encoding these spike mutations (rVSV/SARS-CoV-2<sub>PMS1-1</sub>) exhibited resistance to neutralization by the RU27 plasma panel that was greater in magnitude and consistency than the individual plasma-selected mutants (median 8.0-fold, range 2.7–52.9-fold; Fig. 2e). Nevertheless, 26 of the 27 RU27 plasmas retained residual neutralizing activity against rVSV/SARS-CoV-2<sub>PMS1-1</sub> (Fig. 2e, Extended Data Fig. 6b). We conclude that some neutralizing epitopes are shared among the convalescent antibodies in high-titre plasmas, but neutralizing activity against SARS-CoV-2 is clearly polyclonal and heterogeneous among individuals with respect to epitope targets.

## **Polymutant and variant neutralization**

We generated a panel of HIV-1 pseudotypes bearing the PMS1-1 spike protein, a second PMS protein with a different set of 13 mutations (selected based on similar criteria; known as PMSD4), or naturally occurring variants or relatives of the SARS-CoV-2 spike protein (Extended Data Fig. 7a, b). The panel included several SARS-CoV-2 VOC or VOI spike proteins, and spike proteins from sarbecoviruses found in bats (bCoV-RaTG13), pangolins (pCoV-GD and pCoV-GX) and previously in humans (SARS-CoV), that exhibit varying degrees of sequence divergence from SARS-CoV-2 (refs. 23,24,25) (see Extended Data Fig. 1 for pseudotype characterization). To determine sensitivity and resistance to polyclonal SARS-CoV-2 antibodies, we used an independent set of 21 randomly selected (Ran1–21) convalescent plasmas, and a set of 14 plasmas from recipients who received an mRNA vaccine (Vac1–14), in addition to the RU27 high-titre plasma panel. The PMS spike proteins exhibited a degree of neutralization resistance that fell with the range of that exhibited by the four SARS-CoV-2 VOCs/VOIs and the four other sarbecoviruses (Extended Data Figs. 8a–c, 9a–c). Specifically, PMS1-1 and PMSD4 exhibited neutralization resistance that was greater than B.1.1.7 VOC and B.1.526 VOI, similar to P.1 VOC and less than B.1.351.3 VOC. PMS1-1 and PMSD4 were more resistant to neutralization than pCoV-GD and bCoV-RaTG13, both of which contain a larger number of changes relative to SARS-CoV-2 than the PMS spike proteins. Conversely, the pCoV-GX and SARS-CoV pseudotypes were more resistant to SARS-CoV-2 convalescent or vaccine recipient plasma than PMS1-1 and PMSD4 (Extended Data Figs. 8a–c, 9a–c). Notably, like PMS1-1 and PMSD4, the B.1.351.3 VOC that encodes only nine spike mutations relative to SARS-CoV-2 Wuhan-hu-1 was more resistant to neutralization than sarbecoviruses (pCoV-GD and bCoV-RaTG13) that contain a greater number of substitutions, suggesting that the B.1.351.3 mutations were selected by antibody pressure.

## A neutralization-resistant polymutant

On the basis of the above findings, we attempted to generate a mutant SARS-CoV-2 spike protein that was minimally divergent from SARS-CoV-2 Wuhan-hu-1, yet resistant to neutralization by polyclonal convalescent and vaccine recipient plasma. Successful derivation of such a spike protein would identify a complete list of neutralization epitopes recognized by polyclonal antibodies. We chose 20 naturally occurring mutations, including 8 NTD and 8 RBD changes (Fig. 3a) that either (1) arose in our plasma selection experiments (Fig. 2b), (2) occur in VOCs with reduced neutralization sensitivity (Extended Data Figs. 7, 8), or (3) arose in our previous studies in which human monoclonal antibody resistance was selected<sup>1,2,9</sup>. Naturally occurring deletion mutations in the NTD (Extended Data Fig. 7b), as well as multiple substitutions conferring resistance to classes 1, 2 and 3 RBD-binding antibodies<sup>1,2,9</sup> were included. An rVSV/SARS-CoV-2 derivative encoding the resulting spike

sequence, termed PMS20 (Fig. 3a), was replication competent but attenuated compared with rVSV/SARS-CoV-2<sub>2E1</sub>, suggesting that the 20 mutations confer a fitness cost (Fig. 3b). Nevertheless, HIV-1 pseudotypes bearing PMS20 were similarly infectious to those bearing the parental spike protein (Extended Data Fig. 1) and were highly resistant to neutralization. Indeed, 17 of the 21 random convalescent and 8 of the 14 mRNA vaccine recipient plasmas produced undetectable neutralization of PMS20 pseudotypes (less than 1:50; Fig. 3c). Among the high-titre convalescent RU27 plasmas, 23 of 27 had residual neutralizing activity against PMS20 that was reduced by a median of 32-fold compared with the parental pseudotype (range 2.8–114-fold; Extended Data Fig. 9a). We conclude that the 20 mutations in the PMS20 spike protein are sufficient for evasion of the majority of the antibodies in the plasma of individuals who have been infected by or vaccinated against SARS-CoV-2.

**Fig. 3: Neutralization resistance of polymutant SARS-CoV-2 spike proteins.**

 [figure3](#)

**a**, Design of the PMS20 spike protein with 20 antibody-selected and VOC-associated mutations. **b**, Replication of rVSV/SARS-CoV-2 chimaeras encoding 2E1 (parental) or PMS20 spike proteins in 293T/ACE2cl.22 cells infected at a multiplicity of 0.001 and 0.008, respectively. **c**, Comparative neutralization potency of randomly selected convalescent (Ran1–21; left) and vaccine recipient (Vac1–14; right) plasmas, against Wuhan-hu-1 and PMS20 (**b**) SARS-CoV-2 HIV-1 pseudotypes. For **b** and **c**, the median ± range of two independent determinations is plotted.

[Source data](#)

## Polyclonal neutralization breadth

In contrast to plasmas from individuals who had been infected or vaccinated, a panel of plasmas from 14 individuals, termed ‘infected-then-vaccinated’ (ITV), who had been both infected by SARS-CoV-2 and subsequently received mRNA vaccines<sup>3</sup>

retained neutralizing activity against HIV-1 pseudotypes bearing the PMS20 spike (Fig. 4a, b). Indeed, the PMS20 mutations that reduced Ran21 and Vac14 plasma NT<sub>50</sub> values by a median of 50-fold (range 5.9–225-fold) and 81-fold (range 8.4–229-fold), respectively, caused a median NT<sub>50</sub> reduction of only 18.6-fold (range 3.9–100-fold) for the ITV plasma panel (Fig. 4a, b). Analysis of chimeric SARS-CoV-2/PMS20 spike proteins in which the respective RBDs were exchanged (PMS20(2-RBD) and SARS-CoV-2(PMS-RBD)) indicated that the relative resistance of the PMS20 spike protein to both Ran and ITV plasmas was conferred by multiple spike determinants and that the neutralization breadth in the ITV plasmas was due to antibodies directed at both RBD and non-RBD determinants (Fig. 4a). In addition to the previously reported potent neutralizing activity of ITV plasmas against the B.1.1.7, B.1.525, P.1 and B.1.351.3 VOCs<sup>3</sup>, the ITV plasmas also potently neutralized B.1.617.2 (delta), as well as a recently described variant (A.VOI.V2)<sup>26</sup> that has 11 substitutions and 3 deletions in the spike protein, including an extensively mutated NTD, and is predicted to be resistant to both class 2 and class 3 RBD-binding neutralizing antibodies (Extended Data Fig. 10a).

**Fig. 4: Neutralization breadth of polyclonal antibodies from ITV individuals.**

 figure4

**a**, Comparative neutralization potency (NT<sub>50</sub> values) of random convalescent (Ran1–21; left) and ITV (ITV1–14; right) plasmas against HIV-1 pseudotypes bearing SARS-CoV-2, PMS20 and RBD-exchanged chimeric spike proteins. **b**, Fold difference in NT<sub>50</sub>, comparing neutralization of HIV-1 pseudotypes bearing SARS-CoV-2 and PMS20 spike proteins by Ran1–21, Vac1–15 and ITV1–14 plasmas (the *P* values were calculated using a two-sided Mann–Whitney test; horizontal lines indicate median values). **c**, Neutralization curves for ITV plasmas and the indicated sarbecovirus HIV-1 pseudotypes. For **a** and **c**, median ± range of two independent determinations is plotted.

## Source data

Plasma from the individuals who are ITV also had substantial neutralizing activity against heterologous sarbecovirus HIV-1 pseudotypes, including those that were poorly neutralized by Ran21, Vac14 and RU27 plasma panels and whose RBD and/or NTD sequences are extensively divergent from SARS-CoV-2 (Fig. 4c, Extended Data Fig. 10a,b). The median NT<sub>50</sub> values for the ITV plasmas against sarbecovirus pseudotypes were 5,330 (range 2,369–7,222) for bCoV-RaTG13, 3,617 (range 1,780–6,968) for pCoV-GX, 2,605 (range 1,386–3,181) for bCoV-WIV16 and 1,208 (range 621–2,705) for SARS-CoV (Fig. 4c, Extended Data Fig. 10a). Notably, the neutralizing activity of the ITV plasmas against the divergent sarbecoviruses bCoV-WIV16 and SARS-CoV was similar to that found in the random convalescent plasmas against SARS-CoV-2 Wuhan-hu-1. Thus, the neutralization potency and breadth of polyclonal plasma following mRNA vaccination of individuals who had previously been infected with SARS-CoV-2 seem greater than previously appreciated.

## **Discussion**

These results indicate the presence of abundant neutralizing antibody targets on the SARS-CoV-2 spike protein. Our recent analyses further suggest that affinity maturation, over months of convalescence, confers antibody flexibility and affinity<sup>2,3,27</sup> and can impose a requirement for multiple viral substitutions for escape from individual neutralizing antibodies. Some human monoclonal antibodies thus have substantial activity against SARS-CoV-2 variants and divergent sarbecoviruses<sup>2,28</sup>. Overall, the diversity, maturity and high concentration of neutralizing antibodies probably explains why polyclonal plasma from individuals who have been both infected and subsequently vaccinated could effectively neutralize the otherwise highly neutralization-resistant PMS20 polymutant, as well as sarbecoviruses that are divergent from SARS-CoV-2. While standard mRNA vaccine regimens may be less effective than infection at eliciting individual antibody breadth<sup>29</sup>, it remains to be seen whether polyclonal neutralization potency and breadth can be achieved using appropriately timed boosting with existing SARS-CoV-2 vaccines. If so, existing immunogens may provide robust protection against future SARS-CoV-2 variants, and a degree of protection against diverse potential future sarbecovirus threats. Conversely, PMS proteins encoding numerous neutralization escape mutations may represent useful immunogens to broaden the polyclonal antibody response elicited by first-generation SARS-CoV-2 vaccines.

## **Methods**

### **Cell lines**

293T (American Type Culture Collection (ATCC) CRL-11268), 293T/ACE2cl.22 and HT1080/ACE2.cl.14 cells<sup>5</sup> were derived from original stocks purchased from the ATCC and were assumed to be authenticated by the ATCC. Cells were periodically checked for mycoplasma and retrovirus contamination by DAPI staining and reverse transcriptase assays, respectively.

## SARS-CoV-2 and sarbecovirus spike protein pseudotyped reporter viruses

Plasmids pSARS-CoV-2-SΔ19 and pSARS-CoV-SΔ19 expressing C-terminally truncated SARS-CoV-2 (NC\_045512) and SARS-CoV spike proteins have been previously described<sup>5</sup> and were used to construct the SARS-CoV-2(1-RBD) and SARS-CoV(2-RBD) expression plasmids in which RBD-encoding sequences were reciprocally exchanged. A panel of plasmids expressing spike proteins from SARS-CoV-2 VOC and VOI was constructed in the context of pSARS-CoV-2-SΔ19 (R683G)<sup>5</sup>. Substitutions were introduced using synthetic gene fragments (IDT) or overlap extension PCR-mediated mutagenesis and Gibson assembly. All VOCs/VOIs and polymutant spike proteins also included the R683G substitution, which disrupts the furin cleavage site and generates higher titre virus stocks without significant effects on pseudotyped virus neutralization sensitivity (Extended Data Fig. 1c, d). The potencies with which the plasma neutralized members of the mutant pseudotype panel were compared with potencies against a ‘wild-type’ SARS-CoV-2 spike sequence, carrying R683G where appropriate. Plasmids expressing the spike proteins found in the horseshoe bat (*Rinolophus affinis*) coronavirus bCoV-RaTG13 (ref. 23) as well as the pangolin (*Manis javanica*) coronaviruses from Guangdong, China (pCoV-GD) and Guanxi, China (pCoV-GX)<sup>24,25</sup> were similarly constructed. Spike sequences were codon-modified to maximize homology with the human codon usage optimized of the pSARS-CoV-2 expressing plasmid VG40589-UT (Sinobiological). The 19-amino acid truncated coding sequence (CDS) of bCoV-RaTG13 (QHR63300), pCoV-GD (CoV\_EPI\_ISL\_410721) and pCoV-GX (CoV\_EPI\_ISL\_410542) were synthesized by GeneART and subcloned into pCR3.1 using NheI and XbaI and Gibson assembly, and referred to as pCR3.1-bCoV-RaTG13-SΔ19, pCR3.1-pCoV-GD-SΔ19 and pCR3.1-pCoV-GX-SΔ19, respectively. Pseudotyped HIV-1 particles were generated as previously described<sup>5</sup>. Specifically, virus stocks were harvested 48 h after transfection of 293T cells with pHIV-1 GagPol and pCCNano/LucGFP (Fig. 1) or pNL4-3ΔEnv-nanoluc (all other figures) along with a spike expression plasmid, filtered and stored at -80 °C.

## SARS-CoV-2/sarbecovirus pseudotype neutralization assays

Plasmas were fivefold serially diluted and then incubated with pseudotyped HIV-1 reporter virus for 1 h at 37 °C. The antibody–pseudotype virus mixture was then added to HT1080/ACE2.cl14 cells. After 48 h, cells were washed with PBS, lysed with Luciferase Cell Culture Lysis reagent (Promega), and Nanoluc luciferase activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega) and a Glomax Navigator luminometer (Promega). The relative luminescence units were normalized to those derived from cells infected with the pseudotyped virus in the absence of plasma. The NT<sub>50</sub> was determined using four-parameter non-linear regression (least squares regression method without weighting) (GraphPad Prism).

## Plasma samples

Plasma samples were from individuals who were infected with SARS-CoV-2 for a mean of 1.3 months before plasma donation<sup>7</sup> or from individuals who had received mRNA vaccines at various times before plasma donation<sup>9</sup>. A set of 27 plasma samples from individuals infected with SARS-CoV-2 with high neutralizing activity who had not been vaccinated<sup>7</sup>, termed the ‘RU27’ plasma panel, were used in VSV-SARS-CoV-2 selection procedures, while this panel plus a second set of 21 randomly selected plasmas (selected at random with blinding to neutralization titre or any demographic characteristic) from the same convalescent cohort formed the ‘Ran21’ plasma panel<sup>7</sup>. A set of 14 plasmas donated by individuals who had received a Pfizer/BioNTech mRNA vaccine formed the ‘Vac14’ plasma panel<sup>9</sup>. Finally, a set of 15 plasmas from individuals who were convalescent who had received a Pfizer/BioNTech mRNA vaccine between 6 and 12 months after infection<sup>3</sup> formed the infected-then-vaccinated ‘ITV15’ plasma panel. The study visits and blood draws were obtained with informed consent from all participants under a protocol that was reviewed and approved by the Institutional Review Board of the Rockefeller University (IRB no. DRO-1006, ‘Peripheral Blood of Coronavirus Survivors to Identify Virus-Neutralizing Antibodies’).

## Selection of antibody-resistant rVSV/SARS-CoV-2 variants

To select plasma-resistant spike variants, rVSV/SARS-CoV-2/GFP<sub>1D7</sub> and rVSV/SARS-CoV-2/GFP<sub>2E1</sub> were passaged to generate diversity, and populations containing 10<sup>6</sup> plaque-forming units were incubated with plasma (diluted 1:50–1:400) for 1 h at 37 °C before inoculation of 2 × 10<sup>5</sup> 293T/ACE2cl.22 cells in six-well plates. The following day, the medium was replaced with fresh medium containing the same concentrations of plasma. Supernatant from the wells containing the highest concentrations of plasma antibodies that showed evidence of rVSV/SARS-CoV-2/GFP replication (large numbers of GFP-positive cells or GFP-positive foci) was harvested 24 h later. Thereafter, aliquots (100 µl) of the cleared supernatant from the first

passage (p1) were incubated with the same or increased concentration of plasma and then used to infect  $2 \times 10^5$  293T/ACE2cl.22 cells in six-well plates, as before (p2). In situations where small but expanding GFP-positive foci were observed, the medium was refreshed at 48 h with fresh medium containing no plasma and the virus was harvested 24 h later. We repeated this process for up to six passages or until reduced neutralization potency for the plasma was obvious, as indicated by visual detection of increasing numbers of GFP-positive cells during passage.

To isolate individual mutant viruses by limiting dilution, the selected rVSV/SARS-CoV-2/GFP<sub>1D7</sub> and rVSV/SARS-CoV-2/GFP<sub>2E1</sub> populations were serially diluted in the absence of plasma and aliquots of each dilution added to individual wells of 96-well plates containing  $1 \times 10^4$  293T/ACE2cl.22 cells. Individual viral variants were identified by observing single GFP-positive plaques in individual wells at limiting dilutions. The plaque-purified viruses were expanded, RNA was extracted and spike sequences were determined.

### rVSV/SARS-CoV-2 neutralization assays

Plasma samples were fivefold serially diluted and then incubated with rVSV/SARS-CoV-2/GFP<sub>1D7</sub> and rVSV/SARS-CoV-2/GFP<sub>2E1</sub>, or plaque-purified selected variants thereof, for 1 h at 37 °C. The antibody–recombinant virus mixture was then added to 293T/ACE2.cl22 cells. After 16 h, cells were harvested, and infected cells were quantified by flow cytometry. The percentage of infected cells was normalized to that derived from cells infected with rVSV/SARS-CoV-2 in the absence of plasma. The NT<sub>50</sub> for each plasma was determined using four-parameter non-linear regression (least squares regression method without weighting) (GraphPad Prism).

### Sequence analyses

To identify putative antibody resistance mutations, RNA was isolated from aliquots of supernatant containing selected viral populations or individual plaque-purified variants using NucleoSpin 96 Virus Core Kit (Macherey-Nagel). The purified RNA was subjected to reverse transcription using random hexamer primers and the SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific). The cDNA was amplified using KOD Xtreme Hot Start DNA Polymerase (Millipore Sigma). Specifically, a fragment including the coding region of the extracellular domain of the spike protein was amplified using primers targeting the intergenic region between VSV-M and spike, and the spike intracellular domain. The PCR products were purified and sequenced either using Sanger sequencing or Illumina sequencing as previously described<sup>30</sup>. For Illumina sequencing, 1 μl of diluted DNA was used with 0.25 μl Nextera TDE1 Tagment DNA enzyme (15027865, Illumina) and 1.25 μl TD Tagment

DNA buffer (15027866, Illumina). Then, the DNA was ligated to i5/i7 barcoded primers using the Illumina Nextera XT Index Kit v2 and KAPA HiFi HotStart ReadyMix (2X; KAPA Biosystems). Next, the DNA was purified using AmPure Beads XP (Agencourt), pooled and sequenced (paired end) using Illumina MiSeq Nano 300 V2 cycle kits (Illumina) at a concentration of 12 pM.

For analysis of the Illumina sequencing data, adapter sequences were removed from the raw reads and low-quality reads (Phred quality score of less than 20) using BBduk. Filtered reads were mapped to the codon-optimized SARS-CoV-2 S sequence in rVSV/SARS-CoV-2/GFP and mutations were annotated using Geneious Prime (version 2020.1.2), using a *P* value cut-off of  $10^{-6}$ . RBD-specific variant frequencies, *P* values and read depth were compiled using Python running pandas (1.0.5), numpy (1.18.5) and matplotlib (3.2.2). The parental rVSV/SARS-CoV-2/GFP 2E1 and 1D7 sequences each contain two adaptive mutations (F157S and R685M for 1D7; D215G and R683G for 2E1), but each was considered ‘wild type’ for the purposes of the plasma selection experiments and were subtracted from the analyses of the sequences.

The frequency of amino acid substitutions during rVSV/SARS-CoV-2 passage in plasmas was compared with the frequency of global occurrences of changes at each residue on 11 May 2021 (Los Alamos, COVID-19 Viral Genome Analysis Pipeline; <https://cov.lanl.gov/content/index>)<sup>31</sup>. For comparison of SARS-CoV-2 with sarbecoviruses, amino acid sequences were aligned with Clustal Omega. Using a python script clone of Simplot (<https://jonathanrd.com/20-05-02-writing-a-simplot-clone-in-python/>), the percent identity relative to SARS-CoV-2 was calculated within a rolling window of 100 amino acids, stepping a single residue at a time.

For three-dimensional sliding window analysis of changes in the spike amino acid sequence observed globally and in vitro, the frequency of global occurrences of changes at each residue (Los Alamos, COVID-19 Viral Genome Analysis Pipeline, <https://cov.lanl.gov/content/index>)<sup>31</sup> was divided by the average frequency of change at any residue and projected in the SARS-CoV-2 spike structure PDB 6VXX<sup>32</sup> as relative change frequency using BioStructMap<sup>33,34</sup>. Alternatively, the averaged frequency of substitutions observed after passaging rVSV/SARS-CoV-2 with plasma was divided by the mean substitution frequency and applied as a three-dimensional sliding window over the spike structure. The average frequency of substitutions in a 15 Å radius is represented using a colour spectrum.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Data on plasma-selected mutations are provided in Supplementary Table 1. Neutralization assay datasets generated during the current study are available in the accompanying supplementary source data files. [Source data](#) are provided with this paper.

## Code availability

Analyses used commercially available software. No new code was generated

## References

1. 1. Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).
2. 2. Muecksch, F. et al. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations. *Immunity* **54**, 1853-1868.e7 (2021).
3. 3. Wang, Z. et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 1 year after infection. *Nature* **595**, 426–431 (2021).
4. 4. Sauer, M. M. et al. Structural basis for broad coronavirus neutralization. *Nature Struct. Mol. Biol.* **28**, 478–486 (2021).
5. 5. Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, e20201181 (2020).
6. 6.

Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 (2021).

7. 7.

Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals. *Nature* **584**, 437–442 (2020).

8. 8.

Rogers, T. F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956–963 (2020).

9. 9.

Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).

10. 10.

Liu, L. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).

11. 11.

Cao, Y. et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 1–12 (2020).

12. 12.

McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332-2347.e16 (2021).

13. 13.

Cerutti, G. et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833.e7 (2021).

14. 14.

Fujino, T. et al. Novel SARS-CoV-2 variant identified in travelers from Brazil to Japan. *Emerg. Infect. Dis.* **27**, 1243–1245 (2021).

15. 15.

Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).

16. 16.

Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).

17. 17.

West, A. P. et al. Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York. *Nat. Commun.* **12**, 4886 (2021).

18. 18.

Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).

19. 19.

Tada, T. et al. Decreased neutralization of SARS-CoV-2 global variants by therapeutic anti-spike protein monoclonal antibodies. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.18.431897> (2021).

20. 20.

Andreano, E. et al. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proc. Natl Acad. Sci. USA* **118**, e2103154118 (2021).

21. 21.

Robbiani, D. F. et al. Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico. *Cell* **169**, 597-609.e11 (2017).

22. 22.

Barnes, C. O. et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* **182**, 828–842.e16 (2020).

23. 23.

Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

24. 24.

Lam, T. T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

25. 25.

Zhang, T., Wu, Q. & Zhang, Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* **30**, 1346-1351.e2 (2020).

26. 26.

de Oliveira, T. et al. A novel variant of interest of SARS-CoV-2 with multiple spike mutations detected through travel surveillance in Africa. Preprint at *medRxiv* <https://doi.org/10.1101/2021.03.30.21254323> (2021).

27. 27.

Gaebler, C. et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021).

28. 28.

Tortorici, M. A. et al. Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* **597**, 103–108 (2021).

29. 29.

Cho, A. et al. Anti-SARS-CoV-2 receptor binding domain antibody evolution after mRNA vaccination. *Nature* <https://doi.org/10.1038/s41586-021-04060-7> (2021).

30. 30.

Gaebler, C. et al. Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. *J. Exp. Med.* **216**, 2253–2264 (2019).

31. 31.

Korber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812-827.e19 (2020).

32. 32.

Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281-292.e6 (2020).

33. 33.

Guy, A. J., Irani, V., Richards, J. S. & Ramsland, P. A. BioStructMap: a Python tool for integration of protein structure and sequence-based features. *Bioinformatics* **34**, 3942–3944 (2018).

34. 34.

Guy, A. J. et al. Proteome-wide mapping of immune features onto *Plasmodium* protein three-dimensional structures. *Sci. Rep.* **8**, 4355 (2018).

## Acknowledgements

This work was supported by US National Institutes of Health grants R37AI64003 and R01AI501111 (to P.D.B); R01AI78788 (to T.H.); and P01-AI138398-S1 and 2U19AI111825 (to M.C.N.). D.P. was supported by a Medical Scientist Training Program grant from the NIGMS (T32GM007739) to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program and by the NIAID (F30AI157898). C.G. was supported by the Robert S. Wennett Post-Doctoral Fellowship, in part by the National Center for Advancing Translational Sciences (National Institutes of Health Clinical and Translational Science Award programme, grant UL1 TR001866), and by the Shapiro–Silverberg Fund for the Advancement of Translational Research. P.D.B. and M.C.N. are Howard Hughes Medical Institute Investigators.

## Author information

### Author notes

1. These authors contributed equally: Fabian Schmidt, Yiska Weisblum

## Affiliations

1. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

Fabian Schmidt, Yiska Weisblum, Daniel Poston, Justin DaSilva, Fengwen Zhang, Eva Bednarski, Theodora Hatzioannou & Paul D. Bieniasz

2. Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA

Magdalena Rutkowska, Michel C. Nussenzweig & Paul D. Bieniasz

3. Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA

Alice Cho, Dennis J. Schaefer-Babajew, Christian Gaebler, Marina Caskey & Michel C. Nussenzweig

## Contributions

P.D.B., T.H., M.C.N., FS. and Y.W. conceived, designed and analysed the experiments. F.S. and Y.W. constructed and performed the rVSV/SARS-CoV-2 selection and neutralization experiments. F.S., Y.W., M.R., J.D.S. and E.B. performed the pseudotype neutralization experiments. F.S., T.H. and F.Z. constructed the expression plasmids. A.C. performed next-generation sequencing. D.P. and F.S. performed bioinformatic analysis. M.C., C.G. and D.J.S.-B. executed the clinical protocols, recruited participants and processed samples. P.D.B., T.H., F.S. and Y.W. wrote the manuscript with input from all co-authors.

## Corresponding authors

Correspondence to [Theodora Hatzioannou](#) or [Paul D. Bieniasz](#).

## Ethics declarations

## Competing interests

P.D.B. has received remuneration from Pfizer, Inc. for consulting on mRNA vaccines.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Characterization of HIV-1 pseudotypes bearing the chimeric, mutant, and variant SARS-CoV-2 and sarbecovirus spike proteins.

**a**, Titration of pseudotyped viruses on 293T/ACE2cl.22 cells. Chimeric spike pseudotyped viruses in the upper left panel were built using the unaltered SARS-CoV-2Δ19 and SARS-CoVΔ19 spike protein constructs and a 3-plasmid HIV-1 pseudotyping system (see [Methods](#)). The other panels depict titration of pseudotypes derived using a furin cleavage site mutant SARS-CoV-2Δ19 spike protein (R683G) and a 2-plasmid HIV-1 pseudotyping system (see [Methods](#)). **b**, The same pseudotyped viruses used in (a) were used to infect 3 different 293T/ACE2 clonal cell lines each expressing a different level of ACE2 (MFI = mean fluorescence intensity). **c**, Titration of pseudotypes bearing an unaltered SARS-CoV-2Δ19 spike protein and a furin cleavage site mutant SARS-CoV-2Δ19 spike protein (R683G) generated using a 2-plasmid HIV-1 pseudotype system (see [Methods](#)). **d**, Comparative neutralization potency (NT<sub>50</sub> values) of high titer convalescent (RU27) plasmas against HIV-1 pseudotypes bearing R683G mutant (grey symbols) and unaltered (red symbols) SARS-CoV-2Δ19 spike proteins. For all panels, median ± range of two independent experiments is plotted.

[Source data](#)

### Extended Data Fig. 2 The RU27 SARS-CoV-2 convalescent plasma panel contains neutralizing antibodies targeting RBD and non-RBD determinants.

**a–c**, Correlations of neutralizing potencies of the RU27 plasmas against pseudotypes bearing the indicated pairs of spike proteins. Simple linear regression was used to calculate R<sup>2</sup> and p-values, dashed lines indicate 95% confidence intervals for the regression line.

### Extended Data Fig. 3 Selection pressure on SARS-CoV-2 spike exerted by convalescent plasma.

Frequencies of amino acid substitutions at each codon of the SARS-CoV-2 spike protein following the indicated number of passages (P2-P6) of two independent rVSV-SARS-CoV-2 populations (1D7 and 2E1), in each of the RU27 plasmas, determined by NGS sequencing.

**Extended Data Fig. 4 Neutralization sensitivity of plasma-selected rVSV/SARS-CoV-2 mutants to RU1-20 plasmas.**

Infection, relative to non-neutralized controls, by plaque purified rVSV/SARS-CoV-2 isolates in the presence of the indicated dilutions of the indicated plasmas from the RU27 panel. The same plasmas that were used to select the indicated mutants were used to determine neutralization potency against the respective plaque purified mutants (red) and parental (WT, grey) rVSV/SARS-CoV-2 1D7 or 2E1 viruses. Median ± range of two technical replicates is plotted.

**Extended Data Fig. 5 Neutralization sensitivity of plasma-selected rVSV/SARS-CoV-2 mutants to RU 21-27 plasmas.**

Infection, relative to non-neutralized controls, by plaque purified rVSV/SARS-CoV-2 isolates in the presence of the indicated dilutions of the indicated plasmas from the RU27 panel. The same plasmas that were used to select the indicated mutants were used to determine neutralization potency against the respective plaque purified mutants (red) and parental (WT, grey) rVSV/SARS-CoV-2 1D7 or 2E1 viruses. Median ± range of two technical replicates is plotted.

**Extended Data Fig. 6 Neutralization sensitivity of rVSV/SARS-CoV-2 encoding the PMS1-1 spike.**

**a**, Design of the PMS1-1 polymutant spike protein with 13 plasma-selected spike mutations aggregated in a single spike. **b**, Infection, relative to non-neutralized controls, by rVSV/SARS-CoV<sub>2PMS1-1</sub> (red) and rVSV/SARS-CoV<sub>2E1</sub> (grey) in the presence on the indicated dilutions of the plasmas from the RU27 panel. Median ± range of two technical replicates is plotted.

**Extended Data Fig. 7 Synthetic polymutant and natural variant SARS-CoV-2 spike proteins.**

**a**, Design of the PMS1-1 and PMSD4 polymutant spike proteins with 13 plasma-selected spike mutations aggregated in each spike. **b**, Schematic representation of mutations in naturally occurring VOC/VOI SARS-CoV-2 spike proteins.

**Extended Data Fig. 8 Neutralization potency of random convalescent and vaccine recipient plasmas against polymutant, VOC/VOI, and sarbecovirus HIV-1 pseudotypes.**

**a–c**, Comparative neutralization potency (NT<sub>50</sub> values) of random convalescent (Ran1-21) and vaccine recipient (Vac1-14) plasmas plasma against WT (grey symbols) and the indicated SARS-CoV-2 synthetic polymutant (**a**), natural variant (**b**) or sarbecovirus (**c**) (red symbol) HIV-1 pseudotypes. For all panels, median ± range of two independent experiments is plotted.

[Source data](#)

**Extended Data Fig. 9 Neutralization potency of high titer convalescent plasma against PMS, VOC/VOI, and sarbecovirus HIV-1 pseudotypes.**

**a–c**, Comparative neutralization potency (NT<sub>50</sub> values) of high titer convalescent (RU27) plasma against WT (grey symbols) and indicated polymutant (**a**), SARS-CoV-2 natural variant (**b**) or sarbecovirus (**c**) (red symbol) HIV-1 pseudotypes. For all panels, median ± range of two independent experiments is plotted.

[Source data](#)

**Extended Data Fig. 10 Neutralization potency of plasma from infected-then-vaccinated against VOC/VOI and diverse sarbecovirus HIV-1 pseudotypes.**

**a**, Neutralization potency (NT<sub>50</sub> values) of random convalescent plasmas (grey symbols) or ITV plasmas (red symbols) against SARS-CoV-2 prototype or variant or sarbecovirus HIV-1 pseudotypes. Median of two independent experiments is plotted. Dashed line indicated median NT<sub>50</sub> for random convalescent plasmas against Wuhan-Hu-1 SARS-CoV-2. Numbers above each scatterplot indicate the median NT<sub>50</sub> relative to the median NT<sub>50</sub> for Wuhan-Hu-1 SARS-CoV-2. **b**, Sequence diversity across sarbecovirus spike domains; SARS-CoV-2 and the indicated sarbecovirus spike sequences were aligned with Clustal and compared using Simplot; the percent identity relative to SARS-CoV2 was plotted within a rolling window of 100 amino acids.

## Supplementary information

**Supplementary Table 1**

Substitutions enriched in rVSV/SARS-CoV-2 following selection in neutralizing plasma. Tabulated SARS-CoV-2 spike substitutions and their frequencies in rVSV/SARS-CoV-2 following passage in each of the RU27 plasma samples.

## Reporting Summary

### **Source data**

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 8](#)

[Source Data Extended Data Fig. 9](#)

### **Rights and permissions**

[Reprints and Permissions](#)

### **About this article**

#### **Cite this article**

Schmidt, F., Weisblum, Y., Rutkowska, M. *et al.* High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature* **600**, 512–516 (2021).  
<https://doi.org/10.1038/s41586-021-04005-0>

- Received: 01 July 2021
- Accepted: 07 September 2021
- Published: 20 September 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04005-0>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*Waning COVID super-immunity raises questions about Omicron\*\*](#)

- Max Kozlov

*Nature* (2021)

- [\*\*How bad is Omicron? What scientists know so far\*\*](#)

- Ewen Callaway
  - Heidi Ledford

*Nature* (2021)

- [\*\*COVID super-immunity: one of the pandemic's great puzzles\*\*](#)

- Ewen Callaway

*Nature* (2021)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04005-0>

- Article
- Open Access
- [Published: 07 October 2021](#)

# Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination

- [Alice Cho](#) ORCID: [orcid.org/0000-0001-6354-6148<sup>1 na1</sup>](https://orcid.org/0000-0001-6354-6148),
- [Frauke Muecksch](#) ORCID: [orcid.org/0000-0002-0132-5101<sup>2 na1</sup>](https://orcid.org/0000-0002-0132-5101),
- [Dennis Schaefer-Babajew](#) ORCID: [orcid.org/0000-0002-5380-2950<sup>1 na1</sup>](https://orcid.org/0000-0002-5380-2950),
- [Zijun Wang](#)<sup>1 na1</sup>,
- [Shlomo Finkin](#) ORCID: [orcid.org/0000-0003-4474-2658<sup>1 na1</sup>](https://orcid.org/0000-0003-4474-2658),
- [Christian Gaebler](#) ORCID: [orcid.org/0000-0001-7295-8128<sup>1</sup>](https://orcid.org/0000-0001-7295-8128),
- [Victor Ramos](#) ORCID: [orcid.org/0000-0001-7353-3420<sup>1</sup>](https://orcid.org/0000-0001-7353-3420),
- [Melissa Cipolla](#)<sup>1</sup>,
- [Pilar Mendoza](#)<sup>1</sup>,
- [Marianna Agudelo](#)<sup>1</sup>,
- [Eva Bednarski](#) ORCID: [orcid.org/0000-0003-4438-5590<sup>2</sup>](https://orcid.org/0000-0003-4438-5590),
- [Justin DaSilva](#) ORCID: [orcid.org/0000-0001-6393-0000<sup>2</sup>](https://orcid.org/0000-0001-6393-0000),
- [Irina Shimeliovich](#)<sup>1</sup>,
- [Juan Dizon](#)<sup>1</sup>,
- [Mridushi Daga](#)<sup>1</sup>,
- [Katrina G. Millard](#) ORCID: [orcid.org/0000-0003-2247-3178<sup>1</sup>](https://orcid.org/0000-0003-2247-3178),
- [Martina Turroja](#) ORCID: [orcid.org/0000-0001-9640-0559<sup>1</sup>](https://orcid.org/0000-0001-9640-0559),
- [Fabian Schmidt](#)<sup>2</sup>,
- [Fengwen Zhang](#)<sup>2</sup>,
- [Tarek Ben Tanfous](#)<sup>1</sup>,
- [Mila Jankovic](#)<sup>1</sup>,
- [Thiago Y. Oliveria](#) ORCID: [orcid.org/0000-0002-2654-0879<sup>1</sup>](https://orcid.org/0000-0002-2654-0879),
- [Anna Gazumyan](#)<sup>1</sup>,
- [Marina Caskey](#) ORCID: [orcid.org/0000-0003-1727-8693<sup>1</sup>](https://orcid.org/0000-0003-1727-8693),
- [Paul D. Bieniasz](#) ORCID: [orcid.org/0000-0002-2368-3719<sup>2,3</sup>](https://orcid.org/0000-0002-2368-3719),
- [Theodora Hatzioannou](#) ORCID: [orcid.org/0000-0002-7889-0766<sup>2</sup>](https://orcid.org/0000-0002-7889-0766) &

- [Michel C. Nussenzweig](#)    [ORCID: orcid.org/0000-0003-0592-8564<sup>1,3</sup>](#)

[\*Nature\*](#) volume 600, pages 517–522 (2021)

- 79k Accesses
- 7 Citations
- 2006 Altmetric
- [Metrics details](#)

## Subjects

- [Antibodies](#)
- [SARS-CoV-2](#)
- [Vaccines](#)

## Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection produces B cell responses that continue to evolve for at least a year. During that time, memory B cells express increasingly broad and potent antibodies that are resistant to mutations found in variants of concern<sup>1</sup>. As a result, vaccination of coronavirus disease 2019 (COVID-19) convalescent individuals with currently available mRNA vaccines produces high levels of plasma neutralizing activity against all variants tested<sup>1,2</sup>. Here we examine memory B cell evolution five months after vaccination with either Moderna (mRNA-1273) or Pfizer-BioNTech (BNT162b2) mRNA vaccine in a cohort of SARS-CoV-2-naive individuals. Between prime and boost, memory B cells produce antibodies that evolve increased neutralizing activity, but there is no further increase in potency or breadth thereafter. Instead, memory B cells that emerge five months after vaccination of naive individuals express antibodies that are similar to those that dominate the initial response. While individual memory antibodies selected over time by natural infection have greater potency and breadth than antibodies elicited by vaccination, the overall neutralizing potency of plasma is greater following vaccination. These results suggest that boosting vaccinated individuals with currently available mRNA vaccines will increase plasma neutralizing activity but may not produce antibodies with equivalent breadth to those obtained by vaccinating convalescent individuals.

[Download PDF](#)

# Main

Between 21 January and 20 July 2021, we recruited 32 volunteers with no history of prior SARS-CoV-2 infection receiving either Moderna (mRNA-1273;  $n = 8$ ) or Pfizer-BioNTech (BNT162b2;  $n = 24$ ) mRNA vaccine for sequential blood donation.

Matched samples were obtained at two or three time points. Individuals indicated as ‘prime’ were sampled an average of 2.5 weeks after receiving their first vaccine dose. Individuals who completed their vaccination regimen were sampled an average of 1.3 months after the boost (median = 35.5 days), which is not statistically different from the sampling at 1.3 months in our naturally infected cohort<sup>3</sup> (median = 38.5 days,  $P = 0.21$ ). Individuals sampled at 1.3 months were sampled again approximately 5 months after the second vaccine dose. The volunteers ranged in age from 23 to 78 years old (median = 34.5 years old), 53% were male and 47% were female (for details, see [Methods](#) and Supplementary Tables [1](#) and [2](#)).

## Plasma binding and neutralization assays

Plasma IgM, IgG and IgA responses to SARS-CoV-2 receptor-binding domain (RBD) were measured by enzyme-linked immunosorbent assay (ELISA)<sup>3</sup>. As previously reported by others<sup>2,4,5,6</sup>, there was a significant increase in IgG reactivity to RBD between prime and boost ( $P < 0.0001$ ) (Fig. [1a](#)). IgM and IgA titres were lower than IgG titres and remained low after the second vaccine dose (Extended Data Fig. [1a, b](#)). The magnitude of the response was inversely correlated with age after the prime ( $r = -0.54$ ,  $P = 0.005$ ), but in this limited sample set the age difference was no longer significant at 1.3 or 5 months after the second vaccine dose (Extended Data Fig. [1c, d](#)). Between 1.3 and 5 months after the boost, anti-RBD titres of IgG and IgA decreased significantly. IgG titres decreased by an average of 4.3-fold (range, 1.7- to 10.2-fold), and the loss of activity was directly correlated with the time after vaccination ( $P < 0.0001$ ) (Fig. [1a](#) and Extended Data Fig. [1a, b, e](#)).

**Fig. 1: Plasma ELISAs and neutralizing activity.**

---

 **figure1**

---

**a**, Graph showing area under the curve (AUC) for plasma IgG binding to SARS-CoV-2 RBD after prime and 1.3 and 5 months (m) after the second vaccine dose for  $n = 32$  paired samples. Samples without a prime value are shown in black. **b**, NT<sub>50</sub> values in plasma from pre-pandemic controls (Ctr,  $n = 3$ ), convalescent individuals 1.3 months (ref. 3) and 6.2 months (ref. 7) after infection (grey), and vaccinated individuals ( $n = 32$ ) after prime and 1.3 and 5 months after receiving two doses of mRNA vaccine. Samples without a prime value are shown in black. **c**, NT<sub>50</sub> values (y axis) versus age (x axis) in  $n = 32$  individuals after prime (black) and 1.3 months (orange) or 5 months (green) after boosting with an mRNA vaccine. **d**, Graph showing NT<sub>50</sub> values (y axis) versus days after boost (x axis) in  $n = 32$  individuals receiving two doses of an mRNA vaccine. Samples without a prime value are shown in black. **e**, Plasma neutralizing activity against the indicated SARS-CoV-2 variants of interest/concern ( $n = 15$  paired samples at 1.3 and 5 months after full vaccination). Refer to the [Methods](#) for a list of all substitutions, deletions and insertions in the spike variants. All experiments were performed at least in duplicate. Red bars and values in **a**, **b** and **e** represent geometric mean values. Statistical significance in **a**, **b** and **e** was determined by two-tailed Kruskal–Wallis test with subsequent Dunn's multiple-comparisons test and in **c** and **d** was determined by two-tailed Spearman correlation test.

Neutralizing activity was measured using HIV-1 pseudotyped with the SARS-CoV-2 spike (S) protein<sup>13,7,8</sup>. Naive individuals showed variable responses to the initial vaccine dose, with a geometric mean half-maximal neutralizing titre (NT<sub>50</sub>) of 171 (Fig. 1b and Supplementary Table 2). The magnitude of the neutralizing responses to the initial vaccine dose in naive volunteers was inversely correlated with age ( $r = -0.39$ ,  $P = 0.05$ ) (Fig. 1c). Both binding and neutralizing responses to the second vaccine dose were correlated with the prime responses (binding:  $r = 0.46$ ,  $P = 0.02$

(Extended Data Fig. 1f); neutralizing:  $r = 0.54$ ,  $P = 0.003$  (Extended Data Fig. 1g)), and there was a nearly 12-fold increase in the geometric mean neutralizing response that was similar in men and women with the age-related difference in neutralizing activity eliminated in the individuals in this cohort (Fig. 1b,c, Extended Data Fig. 1h,i). At 1.3 and 5 months after the boost, naive vaccinated individuals had 4.9- and 3.6-fold-higher neutralizing titres, respectively, than seen in a cohort of infected individuals measured at 1.3 months (ref. 3) and 6.2 months (ref. 7) after symptom onset ( $P < 0.0001$ ) (Fig. 1b). Neutralizing responses were directly correlated with anti-RBD IgG titres ( $r = 0.96$ ,  $P < 0.0001$ ) (Extended Data Fig. 1j). Thus, the data obtained from this cohort agree with previous observations showing a significant increase in plasma neutralizing activity that is correlated with improved vaccine efficacy in naive individuals who receive the second dose of mRNA vaccine<sup>2,6,9,10</sup> and higher neutralizing titres in fully vaccinated than in infected individuals<sup>2,6</sup>.

The 28 individuals assayed 5 months after vaccination had a 7.1-fold decrease in geometric mean neutralizing titre relative to their measurement at 1.3 months ( $P < 0.0001$ ) (Fig. 1b), with a range of 1.4- to 27-fold decrease. Neutralizing activity was inversely correlated with the time from vaccination ( $r = -0.82$ ,  $P < 0.0001$ ) (Fig. 1d) and directly correlated with anti-RBD IgG binding titres assessed 5 months after vaccination (Extended Data Fig. 1k). As previously reported by others<sup>11</sup>, the ratio of binding to neutralizing serum titres was significantly higher in vaccinated individuals than in convalescent individuals at the 1.3-month time point ( $P < 0.0001$ ) (Extended Data Fig. 1l). However, a difference was no longer apparent at the later time point (Extended Data Fig. 1l).

It has previously been shown that the neutralizing responses elicited by mRNA vaccination are more potent against the original Wuhan Hu-1 strain than they are against some of the currently circulating variants of concern<sup>2,12,13,14</sup>. To confirm these observations, we measured the neutralizing activity of 15 paired plasma samples obtained from naive individuals 1.3 and 5 months after the second vaccine dose against B.1.1.7 (Alpha variant), B.1.351 (Beta variant), B.1.526 (Iota variant), P.1 (Gamma variant) and B.1.617.2 (Delta variant). Consistent with previous reports<sup>13,15,16,17</sup>, neutralizing activity against the variants was lower than that against the original Wuhan-Hu-1 strain (Fig. 1e and Supplementary Table 3). Initial geometric mean neutralizing titres at 1.3 months against B.1.351, B.1.1.7, B.1.526, P.1 and B.1.617.2 were 5.7-, 1.8-, 1.1-, 1.4- and 2.7-fold lower, respectively, than they were against the Wuhan-Hu-1 virus (Fig. 1e). In the months following vaccination, there was a decrease in neutralizing activity against Wuhan-Hu-1 (R683G) and all the variants, with geometric mean neutralizing titres for wild-type (WT), B.1.351, B.1.1.7, B.1.526, P.1 and B.1.617.2 strains decreasing by 2.9-, 1.8-, 2.3-, 2.9-, 2.4- and 2.6-fold, respectively (Fig. 1e and Supplementary Table 3).

## Monoclonal antibodies

Circulating antibodies produced by plasma cells can prevent infection if present at sufficiently high concentrations at the time of exposure. By contrast, the memory B cell compartment contains long-lived antigen-specific B cells that mediate rapid recall responses that contribute to long-term protection<sup>18</sup>. To examine the nature of the memory compartment elicited by one or two mRNA vaccine doses and its evolution after 5 months, we used flow cytometry to enumerate B cells expressing receptors that bind to Wuhan-Hu-1 (WT) and B.1.351 (K417N/E484K/N501Y) RBDs (Fig. [2a,b](#), and Extended Data Fig. [2](#)). Although neutralizing antibodies develop to other parts of the spike protein, we focused on the RBD because it is the dominant target of the memory antibody neutralizing response<sup>19,20</sup>. Wuhan-Hu-1 RBD-specific memory B cells developed after the prime in all volunteers examined, and their numbers increased for up to 5 months after vaccination (Fig. [2a](#)). Memory B cells binding to the B.1.351 (K417N/E484K/N501Y) variant RBD were detectable but in lower numbers than B cells binding WT RBD in all samples examined (Fig. [2b](#)). Whereas IgG-expressing memory cells increased in number after the boost, IgM-expressing memory B cells that made up 23% of the memory compartment after the prime were nearly absent after boosting (Extended Data Fig. [3a](#)). Finally, circulating RBD-specific plasmablasts were readily detected after the prime but were infrequent after the boost (Extended Data Figs. [2d and 3b](#)).

**Fig. 2: Anti-SARS-CoV-2 RBD B cells after vaccination.**

 [figure2](#)

**a, b,** Graphs summarizing the number of Wuhan-Hu-1 RBD (WT)-specific memory B cells (**a**) and the number of antigen-specific memory B cells cross-reactive with both WT and K417N/E484K/N501Y mutant RBD (**b**) per 10 million B cells for  $n = 32$  individuals after prime and 1.3 and 5 months after full vaccination. Samples without a prime value are shown in black. **c,** Pie charts showing the distribution of IgG antibody sequences obtained for memory B cells from three representative individuals after prime and 1.3 and 5 months after the boost. Additional pie charts can be found in Extended Data Fig. [3](#). The number inside the circle indicates the number of sequences analysed for the individual denoted above the circle, with Pfizer-BioNTech vaccine indicated by (P) and Moderna vaccine indicated by (M). Pie slice size is proportional to the number of clonally related sequences. The black outline and associated numbers

indicate the percentage of clonally expanded sequences detected at each time point. Coloured slices indicate persisting clones (same *IGHV* and *IGLV* genes, with highly similar complementarity-determining region 3 sequences (CDR3s)) found at more than one time point within the same individual, grey slices indicate clones unique to the time point and white slices indicate repeating sequences isolated only once per time point. **d**, Number of nucleotide (nt) somatic hypermutations (SHM) in *IGHV* and *IGLV* genes combined ( $n = 2,050$ ; Supplementary Table 4) in the antibodies illustrated in **c** and Extended Data Fig. 3, compared with the number of mutations obtained 1.3 months (ref. 3) and 6.2 months (ref. 7) after infection (grey). Horizontal bars and red numbers indicate the mean value at each time point. Samples without a prime value are shown in black. Statistical significance in **a**, **b** and **d** was determined by two-tailed Kruskal–Wallis test with subsequent Dunn’s multiple-comparisons test.

The memory compartment continues to evolve up to 1 year after natural infection, with selective enrichment of cells producing broad and potent neutralizing antibodies<sup>1</sup>. To determine how the memory compartment evolves after vaccination, we obtained 2,327 paired antibody sequences from 11 individuals sampled at the time points described above (Fig. 2c, Extended Data Fig. 3c–e and Supplementary Table 4). As expected, *IGHV3-30* and *IGHV3-53* were over-represented after the first and second vaccine dose and remained over-represented 5 months after vaccination<sup>21,22,23</sup> (Extended Data Fig. 4).

All individuals examined showed expanded clones of memory B cells that expressed closely related *IGHV* and *IGHL* genes (Fig. 2c and Extended Data Figs. 3c–e and 4). Paired samples from prime and 1.3 months after the boost showed expanded clones of memory B cells, some of which were shared across plasmablasts, IgM- and IgG-expressing cells at prime, and IgG-expressing memory cells after boost (Extended Data Figs. 3c and 5). Thus, the cell fate decision controlling germinal centre versus plasmablast cell fate is not entirely affinity dependent, as cells with the same initial affinity can enter both compartments to produce clonal relatives<sup>24</sup>.

The relative fraction of memory cells found in expanded clones varied between prime and boost and among individuals and decreased over time (Fig. 2c and Extended Data Fig. 3d–f). Overall, these clones represented 30%, 21% and 9.7% of all sequences after prime and at the 1.3- and 5-month time points, respectively (Extended Data Fig. 3f). Nevertheless, clones of memory B cells continued to evolve for up to 5 months in vaccinated individuals, as evidenced by the appearance of unique clones. Notably, unique clones appearing after 1.3 and 5 months represented a greater or equal fraction of the total memory B cell pool relative to persisting clones (16% versus 9.6% and 5.1% versus 4.7%, respectively) (Fig. 2c and Extended Data Fig. 3d,e,g). Finally, memory B cells emerging after the boost showed significantly higher levels of somatic mutations than plasmablasts or memory B cells isolated after the prime, and they

continued to accumulate mutations up to 5 months after the boost (Fig. 2d and Extended Data Fig. 3h,i). In conclusion, the memory B cell compartment continues to evolve for up to 5 months after mRNA vaccination.

## Neutralizing activity of monoclonal antibodies

We performed ELISAs to confirm that the antibodies isolated from memory B cells bind to the RBD (Extended Data Fig. 6). In total, 458 antibodies were tested by ELISA, including 88 isolated after the first vaccine dose, 210 isolated after the boost and 160 isolated from individuals who had been fully vaccinated 5 months earlier. Among the 458 antibodies tested, 430 (94%) bound to the Wuhan-Hu-1 RBD, indicating that the method used to isolate RBD-specific memory B cells was highly efficient (Supplementary Tables 5–6). The geometric mean ELISA half-maximal effective concentration ( $EC_{50}$ ) of the antibodies obtained after prime and 1.3 and 5 months after the second dose was 3.5, 2.9 and 2.7  $\text{ng ml}^{-1}$ , respectively, suggesting no major change in binding over time after vaccination (Extended Data Fig. 6 and Supplementary Tables 5, 6).

In total, 430 RBD-binding antibodies were tested for neutralizing activity using HIV-1 pseudotyped with the SARS-CoV-2 spike protein<sup>3,8</sup>. The geometric mean half-maximal inhibitory concentration ( $IC_{50}$ ) of RBD-specific memory antibodies improved from 376  $\text{ng ml}^{-1}$  to 153  $\text{ng ml}^{-1}$  between the first and second vaccine dose ( $P = 0.0005$ ) (Fig. 3a). The improvement was reflected in all clones ( $IC_{50} = 377$  versus 171  $\text{ng ml}^{-1}$ ,  $P = 0.01$ ) (Extended Data Fig. 7a), persisting clones ( $IC_{50} = 311$  versus 168  $\text{ng ml}^{-1}$ ) (Fig. 3b and Supplementary Table 6), unique clones ( $IC_{50} = 418$  versus 165  $\text{ng ml}^{-1}$ ,  $P = 0.03$ ) (Fig. 3c) and single antibodies ( $IC_{50} = 374$  versus 136  $\text{ng ml}^{-1}$ ) (Extended Data Fig. 7b). The increase in neutralizing activity between the first and second vaccine doses was associated with a decrease in the percentage of non-neutralizing antibodies (defined as having  $IC_{50} > 1,000 \text{ ng ml}^{-1}$ ) and increased representation of neutralizing antibodies ( $P = 0.003$ ) (Fig. 3a). In conclusion, memory B cells recruited after the second dose account for most of the improvement in neutralizing activity in this compartment when comparing the two vaccine doses. Thus, in addition to the quantitative improvement in serum neutralizing activity, there is also an improvement in the neutralizing activity of the antibodies expressed in the memory compartment after boosting.

**Fig. 3: Anti-SARS-CoV-2 RBD monoclonal antibodies.**

---

 **figure3**

**a–c**, Graphs showing the anti-SARS-CoV-2 neutralizing activity of monoclonal antibodies measured by SARS-CoV-2-pseudotyped virus neutralization assays using WT (Wuhan-Hu-1; ref. [50](#)) SARS-CoV-2 pseudovirus<sup>[3,8](#)</sup>. IC<sub>50</sub> values for all antibodies (**a**), persisting clones (**b**) and unique clones (**c**) isolated from convalescent individuals 1.3 months (ref. [3](#)) and 6.2 months (ref. [7](#)) after infection or from vaccinated individuals after prime and 1.3 and 5 months after the boost are shown. Each dot represents one antibody; 451 total antibodies were tested, including the 430 reported herein (Supplementary Table [5](#)) and 21 previously reported antibodies<sup>[13](#)</sup>. Antibodies isolated from samples without a prime value are shown in black. Pie charts illustrate the fraction of non-neutralizing (IC<sub>50</sub> > 1,000 ng ml<sup>-1</sup>) antibodies (grey slices); the inner circle shows the number of antibodies tested per group. Horizontal bars and red numbers indicate geometric mean values. Statistical significance was determined by two-tailed Kruskal–Wallis test with subsequent Dunn's multiple-comparisons test and for ring plots was determined by two-tailed Fisher's exact test with subsequent Bonferroni correction. All experiments were performed at least twice.

By contrast, there was no significant improvement in neutralizing activity when comparing the monoclonal antibodies obtained 5 months after vaccination with those obtained at 1.3 months ( $P > 0.99$ ) (Fig. [3a](#)). Although there was some improvement among B cell clones, which was accounted for by the small minority of persisting clones, this was not significant in either group ( $P = 0.58$  and  $0.46$ ) (Fig. [3b](#), Extended Data Fig. [7a](#) and Supplementary Table [6](#)). By contrast, memory antibodies obtained from convalescent individuals showed improved neutralizing activity at 6.2 months (ref. [7](#)) relative to 1.3 months (ref. [3](#)), with a decrease in IC<sub>50</sub> from 171 ng ml<sup>-1</sup> to 116 ng ml<sup>-1</sup> (Fig. [3a](#)), and neutralizing activity was further improved after 1 year<sup>[1](#)</sup>. This improvement was due to increased neutralizing activity among persisting clones ( $P = 0.003$ ) (Fig. [3b](#)).

## Affinity, epitopes and neutralization breadth

To examine affinity maturation after vaccination, we performed biolayer interferometry (BLI) experiments using the Wuhan-Hu-1 RBD<sup>3</sup>. In total, 147 antibodies were assayed, 30 obtained after the prime, 74 obtained 1.3 months after boosting and 43 obtained 5 months after boosting. Geometric mean IC<sub>50</sub> values were comparable for the antibodies obtained from the 1.3- and 5-month time points (Extended Data Fig. [8a](#)). Overall, there was a 3- and 7.5-fold increase in affinity for the antibodies obtained between the first two and between the second two time points, respectively (Fig. [4a](#)). After 5 months, the affinity of the antibodies obtained from vaccinated individuals was similar to that for antibodies obtained from naturally infected volunteers (Fig. [4a](#)). However, there was no correlation between the affinity and neutralizing activity of the antibodies tested at any of the three time points (Extended Data Fig. [8b](#)).

**Fig. 4: Affinity and breadth.**

---

 **figure4**

**a, b**, Graphs showing antibody dissociation constant ( $K_d$ ) values for Wuhan-Hu-1 RBD measured by BLI. **a**, Antibodies isolated from convalescent individuals 1.3 months ( $n = 42$ )<sup>3</sup> and 6.2 months ( $n = 45$ )<sup>7</sup> after infection or from vaccinated individuals after prime ( $n = 36$ ) and 1.3 months ( $n = 74$ ) and 5 months ( $n = 43$ ) after the

second vaccination. **b**, Clonally paired antibodies isolated from convalescent individuals 1.3 months (ref. [3](#)) and 6.2 months (ref. [7](#)) after infection ( $n = 15$ ) or vaccinated individuals at prime and 1.3 months ( $n = 3$ ), at prime and 5 months ( $n = 3$ ), or at 1.3 and 5 months after full vaccination ( $n = 26$ ). Antibodies isolated from samples without a prime value are shown in black. Red horizontal bars and numbers indicate median values. **c, d**, Heat maps showing inhibitory concentrations of antibodies isolated 5 months after vaccination (**c**) or 6.2 months (ref. [7](#)) after infection (**d**) normalized to their shared clone isolated 1.3 months after vaccination (**c**) or 1.3 months (ref. [3](#)) after infection (**d**), expressed as % $IC_{50}$ , against the indicated WT or mutant SARS-CoV-2 pseudoviruses (Supplementary Table [8](#)). Antibodies with improved (<30%)  $IC_{50}$  compared with their clonal relative isolated at an earlier time point are shown in shades of green with the most improved antibodies in dark green. Antibodies with worse (>300%)  $IC_{50}$  than their clonal relative isolated at an earlier time point are shown in red with the most worsened antibodies in dark red. Antibodies for which  $IC_{50}$  did not change by more than around 3-fold are shown in yellow. **e**, Pie charts illustrating the fraction of antibodies showing improved (<30%, green) versus not improved (yellow)  $IC_{50}$  values compared with their clonal relative isolated at an earlier time point. The inner circle shows the number of antibody–mutant combinations analysed per group. Statistical significance in **a** and **b** was determined using two-tailed Kruskal–Wallis test with subsequent Dunn’s multiple-comparisons test and in **e** was determined by two-tailed Fisher’s exact test with subsequent Bonferroni correction.

We also compared the affinity for pairs of antibodies obtained from persisting clones at 1.3 and 5 months after vaccination. Persisting clones obtained at 5 months from vaccinated individuals showed a median 4.5-fold increase in affinity relative to the 1.3-month time point ( $P < 0.0001$ ) (Fig. [4b](#)). By contrast, a comparable group of persisting clonal antibodies obtained from convalescent individuals 1.3 and 6.2 months after infection showed a median 11.2-fold increase in affinity at the later time point ( $P = 0.002$ ; Fig. [4b](#)).

To determine whether the epitopes targeted by the monoclonal antibodies were changing over time, we performed BLI experiments in which a preformed antibody–RBD complex was exposed to a second monoclonal antibody targeting one of four classes of structurally defined epitopes<sup>[1,3](#)</sup> (see schematic in Extended Data Fig. [8c](#)). There was no significant change in the distribution of targeted epitopes among 52 randomly selected antibodies, with comparable neutralizing activity obtained at the 1.3- and 5-month time points (Extended Data Figs. [8d, e](#) and [9](#)).

In addition to the increase in potency, the neutralizing breadth of memory antibodies obtained from persisting clones in convalescent individuals increases with time after

infection<sup>1,7,25</sup>. To determine whether there is a similar increase in breadth with time after vaccination, we randomly selected 20 antibodies from prime or 1.3 months after boost with representative levels of activity against the original Wuhan-Hu-1 strain and measured their neutralization potency against a panel of pseudotyped viruses encoding RBD mutations that were selected for resistance to different anti-RBD antibody classes and/or are associated with circulating variants of concern (Extended Data Table 1). There was little change in breadth between prime and 1.3 months after boost, with only a small increase in resistance to variants with the K417N and A475V substitutions (Extended Data Table 1 and Supplementary Table 2).

In addition, we assayed 19 pairs of neutralizing antibodies expressed by persisting clones obtained 1.3 and 5 months after vaccination for their potency against the same mutant pseudotype viruses (Fig. 4c and Supplementary Table 8). These were compared to seven previously reported<sup>25</sup> and nine additional pairs of antibodies obtained from convalescent individuals at 1.3- and 6.2-month time points (Fig. 4d and Supplementary Table 8). Whereas only 36 of 190 (19%) of the antibody–mutant combinations in vaccinated individuals showed improved potency at the later time point, 95 of the 160 (59%) pairs in convalescent individuals exhibited an increase in potency ( $P < 0.0001$ ) (Fig. 4c–e). Moreover, only 4 of the 19 (21%) antibody pairs from vaccinated individuals showed improved potency against pseudotypes carrying B.1.617.2 (Delta variant)-specific RBD amino acid substitutions (L452R/T478K), while 11 of 16 (69%) of the convalescent antibody pairs showed improved activity against this virus ( $P = 0.007$ ) (Fig. 4c–e). We conclude that there is less increase in breadth in the months after mRNA vaccination than there is in a similar interval in naturally infected individuals.

Circulating antibodies are produced by an initial burst of short-lived plasmablasts<sup>26,27</sup> and maintained by plasma cells with variable longevity<sup>28,29</sup>. SARS-CoV-2 infection or mRNA vaccination produces an early peak antibody response that decreases by 5- to 10-fold after 5 months<sup>7,30,31,32,33,34</sup>. Notably, neutralization titres elicited by vaccination exceed those in individuals who have recovered from COVID-19 at all comparable time points assayed. Nevertheless, neutralizing potency against variants is significantly lower than against Wuhan-Hu-1, with up to 5- to 10-fold-reduced activity against the B.1.351 variant<sup>5,6,13,14,35</sup>. Taken together with the overall decay in neutralizing activity, there can be a decrease of 1–2 orders of magnitude in serum neutralizing activity against variants after 5 or 6 months when compared with the peak neutralizing activity against Wuhan-Hu-1. Thus, antibody-mediated protection against variants is expected to wane significantly over a period of months, consistent with reports of re-infection in convalescent individuals and breakthrough infection by variants in fully vaccinated individuals<sup>36,37,38,39</sup>.

In contrast to circulating antibodies, memory B cells are responsible for rapid recall responses<sup>40,41,42</sup>, and the number of cells in this compartment is relatively stable over the first 5–6 months after mRNA vaccination or natural infection<sup>7,43</sup>. In both cases, memory B cells continue to evolve, as evidenced by increasing levels of somatic mutation and emergence of unique clones.

The memory response would be expected to protect individuals who experience breakthrough infection from developing serious disease. Both natural infection and mRNA vaccination produce memory antibodies that evolve increased affinity. However, vaccine-elicited memory monoclonal antibodies show more modest neutralizing potency and breadth than those that develop after natural infection<sup>1,7</sup>. Notably, the difference between the memory compartments that develop in response to natural infection versus mRNA vaccination reported above is consistent with the higher level of protection from variants conferred by natural infection<sup>39</sup>.

There are innumerable differences between natural infection and mRNA vaccination that could account for the differences in antibody evolution over time. These include, but are not limited to, (1) route of antigen delivery (respiratory tract versus intramuscular injection)<sup>44,45</sup>; (2) the physical nature of the antigen (intact virus versus conformationally stabilized pre-fusion spike protein)<sup>46</sup>; and (3) antigen persistence (weeks in the case of natural infection<sup>7</sup> versus hours to days for mRNA vaccination)<sup>47</sup>. Each of these factors could affect B cell evolution and selection directly and indirectly through differential T cell recruitment.

The increase in potency and breadth in the memory compartment that develops after natural infection accounts for the exceptional responses to Wuhan-Hu-1 and its variants that convalescent individuals exhibit when boosted with mRNA vaccines<sup>1,5</sup>. The expanded memory B cell compartment in individuals receiving mRNA vaccines should also produce high titres of neutralizing antibodies when these individuals receive boosts or when they are re-exposed to the virus<sup>48</sup>. Boosting vaccinated individuals with currently available mRNA vaccines should produce strong responses that mirror or exceed the initial vaccine responses to Wuhan-Hu-1, but with similarly decreased coverage against variants. Whether an additional boost with Wuhan-Hu-1-based or variant vaccines or re-infection will also elicit development of memory B cells expressing antibodies showing increased breadth remains to be determined. Finally, timing an additional boost for optimal responses depends on whether the objective is to prevent infection or disease<sup>49</sup>. Given the current rapid emergence of SARS-CoV-2 variants, boosting to prevent infection would probably be needed on a timescale of months. The optimal timing for boosting to prevent serious disease will depend on the stability and further evolution of the memory B cell compartment.

# Methods

## Study participants

Participants were healthy volunteers receiving either the Moderna (mRNA-1273) or Pfizer-BioNTech (BNT162b2) mRNA vaccine against SARS-CoV-2 who were recruited for serial blood donations at Rockefeller University Hospital in New York between 21 January and 20 July 2021. The majority of participants ( $n=28$ ) were de novo recruited for this study, while a subgroup of individuals ( $n=4$ ) were from a long-term study cohort<sup>13</sup>. Eligible participants were healthy adults with no history of infection with SARS-CoV-2, as determined by clinical history and confirmed through serology testing, receiving one of the two Moderna (mRNA-1273) or Pfizer-BioNTech (BNT162b2) vaccines according to current dosing and interval guidelines. Exclusion criteria included incomplete vaccination status, presence of clinical signs and symptoms suggestive of acute infection with SARS-CoV-2, a positive RT–PCR result for SARS-CoV-2 in saliva or positive COVID-19 serology. Seronegativity for COVID-19 was established through the absence of serological activity towards the nucleocapsid (N) protein of SARS-CoV-2. Participants presented to the Rockefeller University Hospital for blood sample collection and were asked to provide details of their vaccination regimen, possible side effects, comorbidities and possible COVID-19 history. Clinical data collection and management were carried out using the software iRIS by iMedRIS (v.11.02). All participants provided written informed consent before participation in the study, and the study was conducted in accordance with Good Clinical Practice principles. The study was performed in compliance with all relevant ethical regulations, and the protocol (DRO-1006) for studies with human participants was approved by the institutional review board of The Rockefeller University. For detailed participant characteristics, see Supplementary Tables 1 and 2.

## Blood sample processing and storage

Peripheral blood mononuclear cells (PBMCs) obtained from samples collected at Rockefeller University were purified as previously reported by gradient centrifugation and stored in liquid nitrogen in the presence of foetal calf serum (FCS) and DMSO<sup>3,7</sup>. Heparinized plasma and serum samples were aliquotted and stored at  $-20^{\circ}\text{C}$  or below. Before experiments, aliquots of plasma samples were heat inactivated ( $56^{\circ}\text{C}$  for 1 h) and then stored at  $4^{\circ}\text{C}$ .

## ELISAs

ELISAs<sup>51,52</sup> to evaluate antibodies binding to SARS-CoV-2 RBD were performed by coating high-binding 96-half-well plates (Corning, 3690) with  $50\ \mu\text{l}$  per well of a  $1\ \mu\text{g}$

$\text{ml}^{-1}$  protein solution in PBS overnight at 4 °C. Plates were washed six times with washing buffer (1× PBS with 0.05% Tween-20 (Sigma-Aldrich)) and incubated with 170  $\mu\text{l}$  per well of blocking buffer (1× PBS with 2% BSA and 0.05% Tween-20 (Sigma)) for 1 h at room temperature. Immediately after blocking, monoclonal antibodies or plasma samples were added in PBS and plates were reincubated for 1 h at room temperature. Plasma samples were assayed at a 1:66 starting dilution with 10 additional threefold serial dilutions. Monoclonal antibodies were tested at a 10  $\mu\text{g ml}^{-1}$  starting concentration with 10 additional fourfold serial dilutions. Plates were washed six times with washing buffer and then incubated with anti-human IgG, IgM or IgA secondary antibody conjugated to horseradish peroxidase (HRP) (Jackson Immuno Research, 109-036-088 and 109-035-129; Sigma, A0295) in blocking buffer at a 1:5,000 dilution (IgM and IgG) or a 1:3,000 dilution (IgA). Plates were developed by addition of the HRP substrate 3,3',5,5'-tetramethylbenzidine (TMB) (ThermoFisher) for 10 min (plasma samples) or 4 min (monoclonal antibodies). The developing reaction was stopped by adding 50  $\mu\text{l}$  of 1 M  $\text{H}_2\text{SO}_4$ , and absorbance was measured at 450 nm with an ELISA microplate reader (FluoStar Omega, BMG Labtech) with Omega and Omega MARS software for analysis. For plasma samples, a positive control (plasma from participant COV72, diluted 66.6-fold with 10 additional threefold serial dilutions in PBS) was added to every assay plate for normalization. The average of its signal was used for normalization of all other values on the same plate with Excel software before calculating the AUC using Prism v9.1(GraphPad). Negative controls of pre-pandemic plasma samples from healthy donors were used for validation (for more details, see ref. <sup>3</sup>). For monoclonal antibodies, the ELISA EC<sub>50</sub> was determined using four-parameter nonlinear regression (GraphPad Prism v9.1). EC<sub>50</sub> values above 2,000 ng  $\text{ml}^{-1}$  were considered to correspond to non-binders.

## Proteins

The mammalian expression vector encoding the RBD of SARS-CoV-2 (GenBank [MN985325.1](#); spike protein residues 319–539) was previously described<sup>53</sup>.

## SARS-CoV-2-pseudotyped reporter virus

The panel of plasmids expressing RBD-mutant SARS-CoV-2 spike proteins in the context of pSARS-CoV-2-S<sub>Δ19</sub> has been described<sup>13,25,54</sup>. Variant pseudoviruses resembling variants of interest/concern B.1.1.7 (first isolated in the UK), B.1.351 (first isolated in South Africa), B.1.526 (first isolated in New York), P.1 (first isolated in Brazil) and B.1.617.2 (first isolated in India) were generated by introduction of substitutions using synthetic gene fragments (IDT) or overlap extension PCR-mediated mutagenesis and Gibson assembly. Specifically, the variant-specific deletions and substitutions introduced were as follows: B.1.1.7: ΔH69/V70, ΔY144,

N501Y, A470D, D614G, P681H, T761I, S982A, D118H; B.1.351: D80A, D215G, L242H, R246I, K417N, E484K, N501Y, D614G, A701V; B.1.526: L5F, T95I, D253G, E484K, D614G, A701V; P.1: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1167F; B.1.617.2: T19R, Δ156–158, L452R, T478K, D614G, P681R, D950N.

The E484K, K417N/E484K/N501Y, L452R/E484Q and L452R/T478K substitutions, as well as the deletions/substitutions corresponding to the variants of concern listed above, were incorporated into a spike protein that also included the R683G substitution, which disrupts the furin cleavage site and increases particle infectivity. Neutralizing activity against mutant pseudoviruses was compared to that against a WT SARS-CoV-2 spike sequence ([NC\\_045512](#)), carrying R683G where appropriate.

SARS-CoV-2-pseudotyped particles were generated as previously described<sup>3,8</sup>. In brief, 293T (CRL-11268) and HT1080 (CCL-121) cells were obtained from ATCC. Cells were transfected with pNL4-3ΔEnv-nanoluc and pSARS-CoV-2-S<sub>Δ19</sub> particles were collected 48 h after transfection, filtered and stored at –80 °C to propagate 293T/ACE2 and HT1080/ACE2.cl14 cells. Cell lines were checked for mycoplasma contamination by Hoechst staining and confirmed negative.

## Pseudotyped virus neutralization assays

Fourfold serially diluted pre-pandemic negative-control plasma from healthy donors, plasma from COVID-19-convalescent individuals or monoclonal antibodies were incubated with SARS-CoV-2-pseudotyped virus for 1 h at 37 °C. The mixture was subsequently incubated with 293T/ACE2 cells<sup>3</sup> (for all WT neutralization assays) or HT1080/ACE2.cl14 cells (for all mutant panels and variant neutralization assays)<sup>13</sup> for 48 h, after which cells were washed with PBS and lysed with Luciferase Cell Culture Lysis 5× reagent (Promega). Nanoluc luciferase activity in lysates was measured using the Nano-Glo Luciferase Assay System (Promega) with the Glomax Navigator (Promega). Relative luminescence units were normalized to those derived from cells infected with SARS-CoV-2-pseudotyped virus in the absence of plasma or monoclonal antibodies. The NT<sub>50</sub> values for plasma or IC<sub>50</sub> and 90% inhibitory concentrations for monoclonal antibodies were determined using four-parameter nonlinear regression (least-squares regression method without weighting; constraints: top = 1, bottom = 0) (GraphPad Prism).

## Biotinylation of viral protein for use in flow cytometry

Purified and Avi-tagged SARS-CoV-2 RBD or SARS-CoV-2 RBD K417N/E484K/N501Y mutant was biotinylated using the Biotin-Protein Ligase-BIRA kit according to the manufacturer's instructions (Avidity) as described before<sup>3</sup>.

Ovalbumin (Sigma, A5503-1G) was biotinylated using the EZ-Link Sulfo-NHS-LC-Biotinylation kit according to the manufacturer's instructions (Thermo Scientific). Biotinylated ovalbumin was conjugated to streptavidin-BV711 (BD Biosciences, 563262), and RBD was conjugated to streptavidin-PE (BD Biosciences, 554061) and streptavidin-AF647 (BioLegend, 405237)<sup>3</sup>.

## Flow cytometry and single-cell sorting

Single-cell sorting by flow cytometry was described previously<sup>3</sup>. In brief, PBMCs were enriched for B cells by negative selection using a pan-B cell isolation kit according to the manufacturer's instructions (Miltenyi Biotec, 130-101-638). The enriched B cells were incubated in FACS buffer (1× PBS, 2% FCS, 1 mM EDTA) with the anti-human antibodies (all at a 1:200 dilution) anti-CD20-PECy7 (BD Biosciences, 335793), anti-CD3-APC-eFluro 780 (Invitrogen, 47-0037-41), anti-CD8-APC-eFluor 780 (Invitrogen, 47-0086-42), anti-CD16-APC-eFluor 780 (Invitrogen, 47-0168-41) and anti-CD14-APC-eFluor 780 (Invitrogen, 47-0149-42), as well as Zombie NIR (BioLegend, 423105) and fluorophore-labelled RBD and ovalbumin (Ova) for 30 min on ice. Single CD3<sup>-</sup>CD8<sup>-</sup>CD14<sup>-</sup>CD16<sup>-</sup>CD20<sup>+</sup>Ova<sup>-</sup>RBD-PE<sup>+</sup>RBD-AF647<sup>+</sup> B cells were sorted into individual wells of 96-well plates containing 4 µl of lysis buffer (0.5× PBS, 10 mM dithiothreitol, 3,000 U ml<sup>-1</sup> RNasin Ribonuclease Inhibitors (Promega, N2615)) per well using a FACSAria III and FACSDiva software (Becton Dickinson) for acquisition and FlowJo software for analysis. The sorted cells were frozen on dry ice and then stored at -80 °C or immediately used for subsequent RNA reverse transcription. For plasmablast single-cell sorting, in addition to the above antibodies, B cells were also stained with anti-CD19-BV605 (BioLegend, 302244) and single CD3<sup>-</sup>CD8<sup>-</sup>CD14<sup>-</sup>CD16<sup>-</sup>CD19<sup>+</sup>CD20<sup>-</sup>Ova<sup>-</sup>RBD-PE<sup>+</sup>RBD-AF647<sup>+</sup> plasmablasts were sorted as described above. For B cell phenotype analysis, in addition to the above antibodies, B cells were also stained with the following anti-human antibodies (all at a 1:200 dilution): anti-IgD-BV421 (BioLegend, 348226), anti-CD27-FITC (BD Biosciences, 555440), anti-CD19-BV605 (BioLegend, 302244), anti-CD71-PerCP-Cy5.5 (BioLegend, 334114), anti-IgG-PECF594 (BD Biosciences, 562538), anti-IgM-AF700 (BioLegend, 314538) and anti-IgA-Viogreen (Miltenyi Biotec, 130-113-481).

## Antibody sequencing, cloning and expression

Antibodies were identified and sequenced as described previously<sup>3,55</sup>. In brief, RNA from single cells was reverse transcribed (SuperScript III Reverse Transcriptase, Invitrogen, 18080-044), and the cDNA was stored at -20 °C or used for subsequent amplification of the variable *IGH*, *IGL* and *IGK* genes by nested PCR and Sanger sequencing. Sequence analysis was performed using MacVector. Amplicons from the first PCR reaction were used as templates for sequence- and ligation-independent

cloning into antibody expression vectors. Recombinant monoclonal antibodies were produced and purified as previously described<sup>3</sup>.

## Biolayer interferometry

BLI assays were performed as previously described<sup>3</sup>. In brief, we used the Octet Red instrument (ForteBio) at 30 °C with shaking at 1,000 r.p.m. Affinity measurement of anti-SARS-CoV-2 IgG binding was corrected by subtracting the signal obtained from traces performed with IgGs in the absence of WT RBD. Kinetic analysis using protein A biosensor (ForteBio, 18-5010) was performed as follows: (1) baseline: immersion for 60 s in buffer; (2) loading: immersion for 200 s in a solution with IgGs at 10 µg ml<sup>-1</sup>; (3) baseline: immersion for 200 s in buffer; (4) association: immersion for 300 s in solution with WT RBD at 20, 10 or 5 µg ml<sup>-1</sup>; (5) dissociation: immersion for 600 s in buffer. Curve fitting was performed using a fast 1:1 binding model and the data analysis software from ForteBio. Mean equilibrium dissociation constants ( $K_d$ ) were determined by averaging all binding curves that matched the theoretical fit with an  $R^2$  value  $\geq 0.8$ .

## Computational analyses of antibody sequences

Antibody sequences were trimmed on the basis of quality and annotated using Igblastn v.1.14 with the IMGT domain delineation system. Annotation was performed systematically using Change-O toolkit v.0.4.540 (ref. <sup>56</sup>). Heavy and light chains derived from the same cell were paired, and clonotypes were assigned on the basis of their V and J genes using in-house R and Perl scripts. All scripts and the data used to process antibody sequences are publicly available on GitHub ([https://github.com/stratust/igipeline/tree/igipeline2\\_timepoint\\_v2](https://github.com/stratust/igipeline/tree/igipeline2_timepoint_v2)).

The frequency distributions of human V genes in anti-SARS-CoV-2 antibodies from this study were compared with 131,284,220 IgH and IgL sequences generated in ref. <sup>57</sup> and downloaded from cAb-Rep<sup>58</sup>, a database of shared human B cell antigen receptor (BCR) clonotypes available at <https://cab-rep.c2b2.columbia.edu/>. On the basis of the 112 distinct V genes that made up the 7,936 analysed sequences from the immunoglobulin repertoire of the 11 participants present in this study, we selected the IgH and IgL sequences from the database that were partially encoded by the same V genes and counted them according to the constant region. The frequencies shown in Extended Data Fig. 4 are relative to the source and isotype analysed. We used the two-sided binomial test to check whether the number of sequences belonging to a specific *IGHV* or *IGLV* gene in the repertoire was different according to the frequency of the same IgV gene in the database. Adjusted  $P$  values were calculated using the false discovery rate (FDR) correction. Significant differences are denoted with asterisks.

Nucleotide somatic hypermutation and CDR3 length were determined using in-house R and Perl scripts. For somatic hypermutations, *IGHV* and *IGLV* nucleotide sequences were aligned against the closest germline sequences using Igblastn and the number of differences was considered to correspond to nucleotide mutations. The average number of mutations for V genes was calculated by dividing the sum of all nucleotide mutations across all participants by the number of sequences used for the analysis.

## Data presentation

Figures were arranged in Adobe Illustrator 2020.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Data are provided in Supplementary Tables 1–8. The raw sequencing data and computer scripts associated with Fig. 2 and Extended Data Fig. 3 have been deposited at GitHub ([https://github.com/stratust/igpipeline/tree/igpipeline2\\_timepoint\\_v2](https://github.com/stratust/igpipeline/tree/igpipeline2_timepoint_v2)). This study also uses data from <https://doi.org/10.5061/dryad.35ks2>, the Protein Data Bank ([6VYB](#) and [6NB6](#)), cAb-Rep (<https://cab-rep.c2b2.columbia.edu/>), the Sequence Read Archive (accession [SRP010970](#)) and ref. <sup>57</sup> (<https://doi.org/10.1038/s41586-019-0934-8>).

## Code availability

Computer code to process the antibody sequences is available at GitHub ([https://github.com/stratust/igpipeline/tree/igpipeline2\\_timepoint\\_v2](https://github.com/stratust/igpipeline/tree/igpipeline2_timepoint_v2)).

## References

1. 1.

Wang, Z. et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021).

2. 2.

Goel, R. R. et al. Distinct antibody and memory B cell responses in SARS-CoV-2 naive and recovered individuals following mRNA vaccination. *Sci. Immunol.* **6**, <https://doi.org/10.1126/sciimmunol.abi6950> (2021).

3. 3.

Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

4. 4.

Apostolidis, S. A. et al. Altered cellular and humoral immune responses following SARS-CoV-2 mRNA vaccination in patients with multiple sclerosis on anti-CD20 therapy. Preprint at <https://doi.org/10.1101/2021.06.23.21259389> (2021).

5. 5.

Sokal, A. et al. Memory B cells control SARS-CoV-2 variants upon mRNA vaccination of naive and COVID-19 recovered individuals. Preprint at <https://doi.org/10.1101/2021.06.17.448459> (2021).

6. 6.

Turner, J. S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021).

7. 7.

Gaebler, C. et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021).

8. 8.

Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, e20201181 (2020).

9. 9.

Pilishvili, T. et al. Interim estimates of vaccine effectiveness of Pfizer-BioNTech and Moderna COVID-19 vaccines among health care personnel—33 U.S. sites, January–March 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 753–758 (2021).

10. 10.

Lopez Bernal, J. et al. Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594 (2021).

11. 11.

Amanat, F. et al. SARS-CoV-2 mRNA vaccination induces functionally diverse antibodies to NTD, RBD, and S2. *Cell* **184**, 3936–3948 (2021).

12. 12.

Reynolds, C. J. et al. Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. *Science* **372**, 1418–1423 (2021).

13. 13.

Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).

14. 14.

Stamatatos, L. et al. mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021).

15. 15.

West, A. P. et al. Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York. Preprint at <https://doi.org/10.1101/2021.02.14.431043> (2021).

16. 16.

Edara, V. V. et al. Infection and vaccine-induced neutralizing-antibody responses to the SARS-CoV-2 B.1.617 variants. *N. Engl. J. Med.* **385**, 664–666 (2021).

17. 17.

Planas, D. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280(2021).

18. 18.

Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).

19. 19.

Dugan, H. L. et al. Profiling B cell immunodominance after SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity* **54**, 1290–1303 (2021).

20. 20.

Li, D. et al. In vitro and in vivo functions of SARS-CoV-2 infection-enhancing and neutralizing antibodies. *Cell* **184**, 4203–4219 (2021).

21. 21.

Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).

22. 22.

Kreer, C. et al. Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. *Cell* **182**, 843–854 (2020).

23. 23.

Seydoux, E. et al. Analysis of a SARS-CoV-2-infected individual reveals development of potent neutralizing antibodies with limited somatic mutation. *Immunity* **53**, 98–105 (2020).

24. 24.

Taylor, J. J., Pape, K. A., Steach, H. R. & Jenkins, M. K. Apoptosis and antigen affinity limit effector cell differentiation of a single naive B cell. *Science* **347**, 784–787 (2015).

25. 25.

Muecksch, F. et al. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations. *Immunity* **54**, 1853–1868 (2021).

26. 26.

Li, G. M. et al. Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proc. Natl Acad. Sci. USA* **109**, 9047–9052 (2012).

27. 27.  
Wrammert, J. et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667–671 (2008).
28. 28.  
Amanna, I. J., Carlson, N. E. & Slifka, M. K. Duration of humoral immunity to common viral and vaccine antigens. *N. Engl. J. Med.* **357**, 1903–1915 (2007).
29. 29.  
Halliley, J. L. et al. Long-lived plasma cells are contained within the CD19<sup>-</sup>CD38<sup>hi</sup>CD138<sup>+</sup> subset in human bone marrow. *Immunity* **43**, 132–145 (2015).
30. 30.  
Dan, J. M. et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, <https://doi.org/10.1126/science.abf4063> (2021).
31. 31.  
Sakharkar, M. et al. Prolonged evolution of the human B cell response to SARS-CoV-2 infection. *Sci. Immunol.* **6**, <https://doi.org/10.1126/sciimmunol.abg6916> (2021).
32. 32.  
Widge, A. T. et al. Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. *N. Engl. J. Med.* **384**, 80–82 (2021).
33. 33.  
Wajnberg, A. et al. Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science* **370**, 1227–1230 (2020).
34. 34.  
Pegu, A. et al. Durability of mRNA-1273-induced antibodies against SARS-CoV-2 variants. Preprint at <https://doi.org/10.1101/2021.05.13.444010> (2021).
35. 35.

Chen, R. E. et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).

36. 36.

Abu-Raddad, L. J., Chemaitelly, H., Butt, A. A. & National Study Group for COVID-19 Vaccination. Effectiveness of the BNT162b2 Covid-19 vaccine against the B.1.1.7 and B.1.351 variants. *N. Engl. J. Med.* **385**, 187–189 (2021).

37. 37.

Hacisuleyman, E. et al. Vaccine breakthrough infections with SARS-CoV-2 variants. *N. Engl. J. Med.* **384**, 2212–2218 (2021).

38. 38.

Lumley, S. F. et al. Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N. Engl. J. Med.* **384**, 533–540 (2021).

39. 39.

Gazit, S. et al. Comparing SARS-CoV-2 natural immunity to vaccine-induced immunity: reinfections versus breakthrough infections. Preprint at <https://doi.org/10.1101/2021.08.24.21262415> (2021).

40. 40.

Mesin, L. et al. Restricted clonality and limited germinal center reentry characterize memory B cell reactivation by boosting. *Cell* **180**, 92–106 (2020).

41. 41.

Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. & Jenkins, M. K. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* **331**, 1203–1207 (2011).

42. 42.

Viant, C. et al. Antibody affinity shapes the choice between memory and germinal center B cell fates. *Cell* **183**, 1298–1311 (2020).

43. 43.

Sokal, A. et al. Maturation and persistence of the anti-SARS-CoV-2 memory B cell response. *Cell* **184**, 1201–1213 (2021).

44. 44.

Feng, L. et al. An adenovirus-vectored COVID-19 vaccine confers protection from SARS-CoV-2 challenge in rhesus macaques. *Nat. Commun.* **11**, 4207 (2020).

45. 45.

Hassan, A. O. et al. A single-dose intranasal ChAd vaccine protects upper and lower respiratory tracts against SARS-CoV-2. *Cell* **183**, 169–184 (2020).

46. 46.

Greaney, A. J. et al. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. *Sci. Transl. Med.* **13**, <https://doi.org/10.1126/scitranslmed.abi9915> (2021).

47. 47.

Pardi, N. et al. Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes. *J. Control. Release* **217**, 345–351 (2015).

48. 48.

Wu, K. et al. Preliminary analysis of safety and immunogenicity of a SARS-CoV-2 variant vaccine booster. Preprint at <https://doi.org/10.1101/2021.05.05.21256716> (2021).

49. 49.

Khoury, D. S. et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021).

50. 50.

Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

51. 51.

Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* **26**, 1033–1036 (2020).

52. 52.

Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2020).

53. 53.

Barnes, C. O. et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* **182**, 828–842 (2020).

54. 54.

Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).

55. 55.

Wang, Z. et al. Enhanced SARS-CoV-2 neutralization by dimeric IgA. *Sci. Transl. Med.* **13**, <https://doi.org/10.1126/scitranslmed.abf1555> (2021).

56. 56.

Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

57. 57.

Soto, C. et al. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398–402 (2019).

58. 58.

Guo, Y., Chen, K., Kwong, P. D., Shapiro, L. & Sheng, Z. cAb-Rep: a database of curated antibody repertoires for exploring antibody diversity and predicting antibody prevalence. *Front. Immunol.* **10**, 2365 (2019).

## Acknowledgements

We thank all study participants who devoted time to our research; The Rockefeller University Hospital nursing staff and Clinical Research Support Office and nursing staff; M. Okawa Frank, M. Bergh and R. B. Darnell for SARS-CoV-2 saliva PCR testing; and C. M. Rice, and all members of the laboratory of M.C.N. for helpful discussions, M. Jankovic for laboratory support and K. Gordon for technical assistance with cell-sorting experiments. This work was supported by NIH grants P01-AI138398-S1 and 2U19-AI111825 to M.C.N., R37-AI64003 to P.D.B. and R01-AI78788 to T.H. F.M. is supported by the Bulgari Women & Science Fellowship in COVID-19 Research. C.G. was supported by the Robert S. Wennett Post-Doctoral Fellowship. D.S.-B. and C.G. were supported in part by the National Center for Advancing Translational Sciences (NIH Clinical and Translational Science Award program, grant UL1-TR001866), and C.G. was supported by the Shapiro–Silverberg Fund for the Advancement of Translational Research. P.D.B. and M.C.N. are Howard Hughes Medical Institute Investigators.

## Author information

### Author notes

1. These authors contributed equally: Alice Cho, Frauke Muecksch, Dennis Schaefer-Babajew, Zijun Wang, Shlomo Finkin

### Affiliations

1. Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA

Alice Cho, Dennis Schaefer-Babajew, Zijun Wang, Shlomo Finkin, Christian Gaebler, Victor Ramos, Melissa Cipolla, Pilar Mendoza, Marianna Agudelo, Irina Shimeliovich, Juan Dizon, Mridushi Daga, Katrina G. Millard, Martina Turroja, Tarek Ben Tanfous, Mila Jankovic, Thiago Y. Oliveria, Anna Gazumyan, Marina Caskey & Michel C. Nussenzweig

2. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

Frauke Muecksch, Eva Bednarski, Justin DaSilva, Fabian Schmidt, Fengwen Zhang, Paul D. Bieniasz & Theodora Hatzioannou

3. Howard Hughes Medical Institute, New York, NY, USA

Paul D. Bieniasz & Michel C. Nussenzweig

## Contributions

P.D.B., T.H. and M.C.N. conceived, designed and analysed data from the experiments. M. Caskey and C.G. designed clinical protocols. A.C., F.M., D.S.-B., Z.W., S.F., P.M., M.A., E.B., J. DaSilva, I.S., J. Dizon, F.S., F.Z., T.B.T. and M.J. carried out experiments. A.G. and M. Cipolla produced antibodies. D.S.-B., M.D., M.T., K.G.M., C.G. and M. Caskey recruited participants and executed clinical protocols. T.Y.O. and V.R. performed bioinformatic analysis. A.C., F.M., D.S.-B., Z.W., S.F. and M.C.N. wrote the manuscript with input from all co-authors.

## Corresponding authors

Correspondence to [Marina Caskey](#), [Paul D. Bieniasz](#), [Theodora Hatzioannou](#) or [Michel C. Nussenzweig](#).

## Ethics declarations

## Competing interests

The Rockefeller University has filed a provisional patent application in connection with this work on which M.C.N. is an inventor (US patent 63/021,387). The patent has been licensed by Rockefeller University to Bristol Meyers Squibb.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Plasma ELISA and neutralization.](#)

**a,b**, Graph shows area under the curve (AUC, Y-axis) for plasma IgM (**a**) or IgA (**b**) antibody binding to SARS-CoV-2 RBD after prime, and 1.3- and 5-months post-boost for paired samples from n=32 vaccinated individuals. Samples without a prime value

are shown in black. **c**, Graph shows plasma IgG antibody binding (AUC, Y-axis) plotted against age (X-axis) after prime (black), and 1.3 months (orange) and 5 months (green) post-second vaccination in n=32 vaccinated individuals. **d**, Graph shows age (years, X-axis) vs. fold-change of IgG-binding titers (AUC, Y-Axis) between prime and 1.3m (orange) or 5m (green) post-boost in n=32 vaccinated individuals. **e**, Graph shows plasma IgG antibody binding AUC values (Y-axis) plotted against time after vaccination (day, X-axis) from n=32 vaccinated individuals. Samples without a prime value are shown in black. **f**, IgG antibody binding after prime (AUC, X-axis) vs. IgG antibody binding after 1.3 months post-boost (AUC, Y-axis) (n=26). **g**, NT50 values after prime (X-axis) vs. NT50 values after 1.3 months post-boost (Y-axis) in individuals receiving two doses of an mRNA vaccine (n=26). **h**, NT50 values after prime and 1.3 months post-boost in females and males receiving 2 doses of an mRNA vaccine (n=26).**i**, Graph shows age (years, X-axis) vs fold-change of NT50 (X-axis) between prime and 1.3m (orange) or 5m (green) post-boost (n=26). **j**, NT50 values (Y-axis) vs. IgG antibody binding (AUC, X-axis) 1.3 months after 2 doses of an mRNA vaccine (n=26). **k**, NT50 values (Y-axis) vs. IgG antibody binding (AUC, X-axis) 5 months after boost in individuals receiving two doses of an mRNA vaccine (n=28). **l**, Ratio of anti-RBD IgG antibody (AUC) to NT50 values (Y-axis) plotted for convalescent infected individuals (grey) 1.3m<sup>3</sup> or 6.2m<sup>7</sup> after infection, and from n=32 vaccinated individuals after the prime, and 1.3m and 5m after receiving 2 doses of an mRNA vaccine. Samples without a prime value are shown in black. All experiments were performed at least in duplicate. Red values or bar in **a**, **b**, **h** and **l** represent geometric mean values. Statistical significance in **a**, **b**, **h**, and **l** was determined by two-tailed Kruskal-Wallis test with subsequent Dunn's multiple comparisons, or by two-tailed Spearman correlation test in **c**, **d**, **e**, **f**, **g**, **i**, **j**, and **k**.

## Extended Data Fig. 2 Flow Cytometry.

**a**, Gating strategy for phenotyping. Gating was on singlets that were CD19<sup>+</sup> or CD20<sup>+</sup> and CD3-CD8-CD16-Ova-. Anti-IgG, IgM, IgA, IgD, CD71 and CD27 antibodies were used for B cell phenotype analysis. Antigen-specific cells were detected based on binding to RBD WT-PE<sup>+</sup> and RBD WT/KEN (K417N/E484K/N501Y)-AF647<sup>+</sup>. **b-c**, Flow cytometry plots showing the frequency of **b**, RBD WT-binding memory B cells, and **c**, RBD-binding memory B cells cross-reactive with WT and K417N/E484K/N501Y mutant RBD in 5 selected individuals, after prime, 1.3 months, and 5 months post-second vaccination. **d**, Flow cytometry plots showing frequency of RBD-binding plasmablasts, in 10 selected vaccinees after prime or 1.3 months post-boost. **e**, Gating strategy for single-cell sorting for CD20+ memory B cells (top panel) or CD19+CD20- plasmablasts (bottom panel) which were double positive for RBD-PE and RBD-AF647. **f-g**, Representative flow cytometry plots showing dual AlexaFluor-647-RBD and PE-RBD-binding, single-cell sorted B cells from **f**, 6 individuals after

prime and 1.3 months or 5 months post-boost and **g**, 5 individuals from 1.3- or 5-months post-boost. Percentage of RBD-specific B cells is indicated.

### Extended Data Fig. 3 anti-SARS-CoV-2 RBD-specific plasmablast and memory B cells responses after vaccination.

**a-b**, Graph showing the **a**, frequency of IgM, IgG, or IgA isotype expression by Wuhan-Hu RBD-specific memory B cells after prime or 1.3 months post-boost (n=10), and **b**, number of Wuhan-Hu RBD-binding plasmablasts per 10 million B cells (n=26) after prime or 1.3 months post-boost. Red numbers indicate geometric means. Gating strategy is in Extended Data Fig. 2. **c-e**, Pie charts show the distribution of IgG antibody sequences obtained from **c**, 6 individuals after prime (upper panel) or 1.3 months post-boost (lower panel). Sequences derived from IgG plasmablast (PB), IgM memory B cells (MBC), and IgG MBC compartments were analyzed after prime, while only IgG MBCs were analyzed at 1.3 months after boost, as indicated to the left of the plots. Pie charts showing only IgG memory B cells from 8 individuals (in additional to the 3 vaccinees shown in Fig. 2c) after **d**, prime and 1.3-months post-boost and **e**, 1.3- and 5-months post-boost. The number inside the circle indicates the number of sequences analyzed for the individual denoted above the circle, with Pfizer vaccinees indicated by (P) and Moderna by (M). Pie slice size is proportional to the number of clonally related sequences. The black outline and associated numbers indicate the percentage of clonally expanded sequences detected at each time point. Colored slices indicate persisting clones (same *IGHV* and *IGLV* genes, with highly similar CDR3s) found at more than one timepoint within the same individual. Grey slices indicate clones unique to the timepoint. White slices indicate repeating sequences isolated only once per time point. **f**, Graph shows the relative percentage of clonal sequences of IgG memory B cells at each time point from n=11 vaccinated individuals illustrated in Fig. 2c and Extended Data Fig. 3d,e. The red numbers indicate the geometric means. Samples without a prime value are shown in black. **g**, Graph shows the percentage of total paired-sequences from IgG memory B cells (n=2050) analyzed at either prime, 1.3- or 5-months post-boost, that can be found as part of all clones (black bars), persisting clones (red bars), unique clones (grey bars), or singlets (white bar). **h-i**, Ratio of the number of somatic nucleotide mutations over the nucleotide length of the V gene in the Ig heavy and light chains, separately, in antibodies detected in **h**, different B cell compartments after prime or 1.3 months post-boost (n=1565) and **i**, IgG memory B cells at 1.3 or 5 months post-boost (n=1610) compared to convalescent infected (grey) individuals after 1.3<sup>3</sup> and 6.2<sup>7</sup> months post-infection (also Supplementary Table 4). Horizontal bars and red numbers indicate mean ratio in each compartment at each time point. Sequences derived from samples without a prime value are shown in black. Statistical significance in **a** and **b** was determined using a two-tailed Wilcoxon matched-pairs signed rank test. **f**, **h**, and **i** was

determined by two-tailed Kruskal Wallis test with subsequent Dunn's multiple comparisons.

#### **Extended Data Fig. 4 Frequency distribution of human V genes.**

**a**, Circos plot depicting relationship between antibodies that share V and J gene usage in both IgH and IgL when comparing prime/1.3m IgG MBC sequences. Purple, green, and grey lines connect related clones, clones and singlets, and singlets to each other, respectively. **b**, Graph shows relative abundance of human heavy chain *IGHV* (top), light chain *IGKV* (middle) or *IGLV* (bottom) genes comparing Sequence Read Archive accession SRP010970 (orange), and IgG MBCs after prime (blue) or 1.3 months post-boost (green). Statistical significance was determined by two-sided binomial test. \* =  $p \leq 0.05$ , \*\* =  $p \leq 0.01$ , \*\*\* =  $p \leq 0.001$ , \*\*\*\* =  $p \leq 0.0001$ . Color of stars indicates: black - comparing Database versus Prime; blue - comparing Database versus 1.3m; red - comparing Prime versus 1.3m. **c**, Circos plot depicting relationship between antibodies that share V and J gene usage in both IgH and IgL when comparing 1.3 m/5 m IgG MBC sequences. Purple, green, and grey lines connect related clones, clones and singlets, and singlets to each other, respectively. **d**, Graph shows relative abundance of human heavy chain *IGHV* (top), light chain *IGKV* (middle) or *IGLV* (bottom) genes comparing Sequence Read Archive accession SRP010970 (orange), and IgG MBCs after 1.3 months (blue) or 5 months (green) post-vaccination. Statistical significance was determined by two-sided binomial test. \* =  $p \leq 0.05$ , \*\* =  $p \leq 0.01$ , \*\*\* =  $p \leq 0.001$ , \*\*\*\* =  $p \leq 0.0001$ . Color of stars indicates: black - comparing Database versus 1.3 months; blue - comparing Database versus 5 months; red - comparing 1.3 months versus 5 months.

#### **Extended Data Fig. 5 Somatic hypermutation of anti-SARS-CoV-2 RBD antibody clones after prime or boost.**

Clonal evolution of RBD-binding B cells from 3 individuals for which plasmablasts, IgM memory B cells, and IgG memory B cells were analyzed after prime, and IgG memory B cells were analyzed after 1.3 months post-boost (as described in Extended Data Fig. 3). The number of somatic nucleotide mutations found in shared clonal families found in at least 2 different compartments is graphed to the right of each donut plot. Color of dot plots match the color of pie slices within the donut plot, which indicate persisting clones. nd – clone was Not Detected in the indicated compartment. Black horizontal line indicates median number of SHM.

#### **Extended Data Fig. 6 Anti-SARS-CoV-2 RBD monoclonal antibodies ELISAs.**

**a-e**, Graphs show anti-SARS-CoV-2 binding activity of n=458 monoclonal antibodies measured by ELISA against RBD. ELISA half-maximal concentration ( $EC_{50}$ ) values for all antibodies (a), all clones (b), persisting clones (c), unique clones (d) and singlets (e) isolated from COVID-19 convalescent individuals 1.3<sup>3</sup> and 6.2<sup>7</sup> months after infection (left panel) or from vaccinated individuals after prime, or 1.3m or 5m after receiving the second dose of mRNA vaccination (right panel). Each dot represents one antibody. Antibodies isolated from samples without a prime value are shown in black. Red horizontal bars and numbers indicate geometric mean values. Statistical significance was determined by two-tailed Mann-Whitney test (left panels of a, b, d and e), two-tailed Kruskal-Wallis test with subsequent Dunn's multiple comparisons (right panels of a-e) or by two-tailed Wilcoxon test (left panel of c). All experiments were performed at least twice.

### Extended Data Fig. 7 Anti-SARS-CoV-2 RBD monoclonal antibodies.

**a-c**, Graphs show anti-SARS-CoV-2 neutralizing activity of monoclonal antibodies measured by a SARS-CoV-2 pseudotype virus neutralization assay using wild-type (Wuhan Hu-1<sup>50</sup>) SARS-CoV-2 pseudovirus<sup>3,8</sup>. Half-maximal inhibitory concentration ( $IC_{50}$ ) values for antibodies from a, all clones and e, singlets isolated from COVID-19 convalescent individuals 1.3<sup>3</sup> and 6.2<sup>7</sup> months after infection or from vaccinated individuals after prime, and 1.3- or 5-months after 2 doses of vaccine. Each dot represents one antibody, where 451 total antibodies were tested including the 430 reported herein (Supplementary Table 5), and 21 previously reported antibodies<sup>13</sup>. Antibodies isolated from samples without a prime value are shown in black. Pie charts illustrate the fraction of non-neutralizing ( $IC_{50} > 1000$  ng/ml) antibodies (grey slices), inner circle shows the number of antibodies tested per group. Horizontal bars and red numbers indicate geometric mean values. Statistical significance was determined by two-tailed Kruskal Wallis test with subsequent Dunn's multiple comparisons, and for ring plots by two-tailed Fisher's exact test with subsequent Bonferroni-correction. All experiments were performed at least twice.

### Extended Data Fig. 8 Affinity and Epitope targeting of anti-SARS-CoV-2 RBD antibodies.

**a**,  $IC_{50}$  values for randomly selected antibodies isolated from convalescents 1.3m<sup>3</sup> (n=42) and 6.2m<sup>7</sup> (n=45) after infection or from vaccinees after prime (n=36), and 1.3m (n=74) and 5m (n=43). Red horizontal lines and numbers indicate geometric mean. Antibodies isolated from samples without a prime value are shown in black. **b**, Graphs show affinities ( $K_D$ , Y-axis) plotted against neutralization activity ( $IC_{50}$ , X-axis) for antibodies isolated after prime (black), or 1.3m (orange) or 5m (green) post-boost vaccination for antibodies shown in a. **c**, Schematic representation of the BLI

experiment for randomly selected antibodies isolated from vaccinees 1.3- and 5 months after full vaccination (each presented group shows n=26 antibodies). **d**. Heatmap of relative inhibition of Ab2 binding to the preformed Ab1-RBD complexes (grey=no binding, yellow=low binding, orange=intermediate binding, red=high binding). Values are normalized through the subtraction of the autologous antibody control. BLI traces can be found in Extended Data Fig. 9. **e**. Pie charts indicate the fraction of antibodies that are assigned to different classes according to their binding pattern as shown in **d** and Extended data Fig. 9. Number in inner circle shows number of antibodies tested. Statistical significance was determined using a two-tailed Kruskal Wallis test with subsequent Dunn's multiple comparisons in **a** and two-tailed Spearman correlation test in **b**, and a two-tailed Chi-square test in **e**.

### Extended Data Fig. 9 BLI traces from epitope mapping of anti-SARS-CoV-2 RBD antibodies.

**a, b**, BLI traces from competition experiments used to determine epitope targets of anti-SARS-CoV-2 RBD antibodies isolated from vaccinees at 1.3m (**a**) or 5m (**b**) post-boost, as illustrated in Extended Data Fig. 8.

### **Extended Data Table 1 Breadth of anti-SARS-CoV-2 RBD antibodies elicited after prime and 2 doses of vaccination**

## **Supplementary information**

### Reporting Summary

### Supplementary Table 1

Summary of cohort characteristics.

### Supplementary Table 2

Summary of individual participant characteristics.

### Supplementary Table 3

Half-maximal neutralization titres (NT<sub>50</sub> values) of plasma against different mutant SARS-CoV-2 pseudoviruses, for paired samples from participants 1.3 and 5 months after the second vaccination.

### Supplementary Table 4

Sequences of anti-SARS-CoV-2 RBD antibodies derived from single cells.

### **Supplementary Table 5**

Sequences, RBD binding ( $EC_{50}$ ) and neutralization ( $IC_{50}$ ) of cloned recombinant antibodies.

### **Supplementary Table 6**

RBD binding ( $EC_{50}$ ) and neutralization ( $IC_{50}$ ) of shared clones of recombinant antibodies isolated after vaccination.

### **Supplementary Table 7**

Neutralization activity of recombinant antibodies against mutant SARS-CoV-2 pseudoviruses, in antibodies isolated after the prime and 1.3 months after the boost vaccination.

### **Supplementary Table 8**

Neutralization activity of recombinant antibodies against mutant SARS-CoV-2 pseudoviruses, in clonal pairs of antibodies isolated at 1.3 and 6.2 months after infection, as well as 1.3 and 5 months after the second vaccination.

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### **Reprints and Permissions**

## **About this article**

## Cite this article

Cho, A., Muecksch, F., Schaefer-Babajew, D. *et al.* Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination. *Nature* **600**, 517–522 (2021). <https://doi.org/10.1038/s41586-021-04060-7>

- Received: 25 July 2021
- Accepted: 24 September 2021
- Published: 07 October 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04060-7>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape\*\*](#)

- Fabian Schmidt
- Yiska Weisblum
- Paul D. Bieniasz

*Nature* (2021)

- [\*\*The germinal centre B cell response to SARS-CoV-2\*\*](#)

- Brian J. Laidlaw

- Ali H. Ellebedy

*Nature Reviews Immunology* (2021)

- **Waning COVID super-immunity raises questions about Omicron**

- Max Kozlov

*Nature* (2021)

- **COVID super-immunity: one of the pandemic's great puzzles**

- Ewen Callaway

*Nature* (2021)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04060-7>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 11 October 2021](#)

# Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity

- [Carolina Lucas](#) [ORCID: orcid.org/0000-0003-4590-2756](#)<sup>1 na1</sup>,
- [Chantal B. F. Vogels](#)<sup>2 na1</sup>,
- [Inci Yildirim](#)<sup>3,4 na1</sup>,
- [Jessica E. Rothman](#) [ORCID: orcid.org/0000-0002-9961-1092](#)<sup>2</sup>,
- [Peiwen Lu](#) [ORCID: orcid.org/0000-0001-6118-872X](#)<sup>1</sup>,
- [Valter Monteiro](#)<sup>1</sup>,
- [Jeff R. Gehlhausen](#)<sup>1,5</sup>,
- [Melissa Campbell](#)<sup>6</sup>,
- [Julio Silva](#) [ORCID: orcid.org/0000-0001-8212-7440](#)<sup>1</sup>,
- [Alexandra Tabachnikova](#)<sup>1</sup>,
- [Mario A. Peña-Hernandez](#)<sup>1</sup>,
- [M. Catherine Muenker](#) [ORCID: orcid.org/0000-0002-0450-0868](#)<sup>2</sup>,
- [Mallery I. Breban](#) [ORCID: orcid.org/0000-0002-8976-9988](#)<sup>2</sup>,
- [Joseph R. Fauver](#) [ORCID: orcid.org/0000-0001-8099-2991](#)<sup>2</sup>,
- [Subhasis Mohanty](#)<sup>1,6</sup>,
- [Jiefang Huang](#)<sup>1,6</sup>,
- [Yale SARS-CoV-2 Genomic Surveillance Initiative](#),
- [Albert C. Shaw](#)<sup>1,6</sup>,
- [Albert I. Ko](#) [ORCID: orcid.org/0000-0001-9023-2339](#)<sup>2,6</sup>,
- [Saad B. Omer](#) [ORCID: orcid.org/0000-0002-5383-3474](#)<sup>2,4,6 na2</sup>,
- [Nathan D. Grubaugh](#)<sup>2,7 na2</sup> &
- [Akiko Iwasaki](#) [ORCID: orcid.org/0000-0002-7824-9856](#)<sup>1,8 na2</sup>

*Nature* volume 600, pages 523–529 (2021)

- 73k Accesses
- 7 Citations

- 1276 Altmetric
- [Metrics details](#)

## Subjects

- [Antibodies](#)
- [RNA vaccines](#)
- [SARS-CoV-2](#)

## Abstract

The emergence of SARS-CoV-2 variants with mutations in major neutralizing antibody-binding sites can affect humoral immunity induced by infection or vaccination<sup>1,2,3,4,5,6</sup>. Here we analysed the development of anti-SARS-CoV-2 antibody and T cell responses in individuals who were previously infected (recovered) or uninfected (naive) and received mRNA vaccines to SARS-CoV-2. While individuals who were previously infected sustained higher antibody titres than individuals who were uninfected post-vaccination, the latter reached comparable levels of neutralization responses to the ancestral strain after the second vaccine dose. T cell activation markers measured upon spike or nucleocapsid peptide in vitro stimulation showed a progressive increase after vaccination. Comprehensive analysis of plasma neutralization using 16 authentic isolates of distinct locally circulating SARS-CoV-2 variants revealed a range of reduction in the neutralization capacity associated with specific mutations in the spike gene: lineages with E484K and N501Y/T (for example, B.1.351 and P.1) had the greatest reduction, followed by lineages with L452R (for example, B.1.617.2). While both groups retained neutralization capacity against all variants, plasma from individuals who were previously infected and vaccinated displayed overall better neutralization capacity than plasma from individuals who were uninfected and also received two vaccine doses, pointing to vaccine boosters as a relevant future strategy to alleviate the effect of emerging variants on antibody neutralizing activity.

[Download PDF](#)

## Main

The ongoing evolution and emergence of SARS-CoV-2 variants raise concerns about the effectiveness of monoclonal antibody therapies and vaccines. The mRNA-based vaccines Pfizer-BioNTech BNT162b2 and Moderna mRNA-1273 encode a stabilized full-length SARS-CoV-2 spike ectodomain derived from the Wuhan-Hu-1 genetic

sequence and elicit potent neutralizing antibodies (NAbs)<sup>7,8</sup>. However, emerging SARS-CoV-2 variants with mutations in the spike gene, especially in NAb-binding sites, have been associated with increased transmissibility<sup>9,10</sup> as well as neutralization resistance to monoclonal antibodies, convalescent plasma and sera from vaccinated individuals<sup>1,2,3,4,5,6</sup>. To better understand how immune responses triggered by vaccination and/or SARS-CoV-2 infection fare against emerging virus variants, we assembled a cohort of mRNA-vaccinated individuals, previously infected or not, and characterized virus-specific immunological profiles. We examined the effect of SARS-CoV-2 variants containing many different key spike gene mutations in mRNA-vaccinated individuals using a comprehensive set of full-length authentic SARS-CoV-2 isolates.

## Vaccine-induced antibody responses

First, to characterize SARS-CoV-2-specific adaptive immune responses after mRNA COVID-19 vaccines (Moderna or Pfizer), 40 healthcare workers (HCWs) from the Yale-New Haven Hospital were enrolled in this study between November 2020 and January 2021, with a total of 198 samples. We stratified the vaccinated participants based on previous exposure to SARS-CoV-2 into previously infected (recovered) or uninfected (naive) groups. Previous infection was confirmed by reverse-transcription quantitative PCR (RT-qPCR) and SARS-CoV-2 IgG enzyme-linked immunosorbent assay (ELISA). The HCWs received mRNA vaccines, either Pfizer or Moderna, and we followed them longitudinally pre-vaccination and post-vaccination (Fig. 1a). Cohort demographics, vaccination status and serostatus are summarized in Extended Data Table 1. We collected plasma and peripheral blood mononuclear cells (PBMCs) sequentially in five time points that covered a period of 98 days after the first vaccination dose. Samples were collected at baseline (before vaccination), 7 and 28 days after the first vaccination dose, and 7, 28 and 70 days after the second vaccination dose (Fig. 1a). We determined antibody profiles, using both ELISA and neutralization assays, and assessed cellular immune response, profiled by flow cytometry using frozen PBMCs.

**Fig. 1: Temporal dynamics of anti-SARS-CoV-2 antibodies in vaccinated participants.**

---

 **figure1**

---

**a**, Cohort timeline overview indicated by days after SARS-CoV-2 mRNA vaccination. HCW participants received two doses of the mRNA vaccine and plasma samples were collected as indicated. Baseline (time point 0 (TP 0)), before vaccination; TP 1, 7 days after the first dose; TP 2, 28 days after the first dose; TP 3, 7 days after the second dose; TP 4, 28 days after the second dose; TP 5, 70 days after the second dose. Participants were stratified on the basis of previous exposure to SARS-CoV-2 (purple: vaccinated, uninfected; blue: vaccinated, previously infected). Created with BioRender.com. **b, c**, Plasma reactivity to the spike (S) protein, receptor-binding domain (RBD) and nucleocapsid (N) measured over time by ELISA. The levels of anti-spike, anti-spike subunit 1 (S1), anti-RBD and anti-nucleocapsid IgG are shown (**b**). A comparison of the levels of anti-spike, anti-S1, anti-RBD and anti-nucleocapsid IgG in vaccinated participants previously infected or not to SARS-CoV-2 is also displayed (**c**). **d, e**, Longitudinal neutralization assay using wild-type SARS-CoV-2, ancestral strain (WA1, USA). The neutralization titre (PRNT50) over time (**d**), and the plasma neutralization capacity between vaccinated participants previously infected or not to SARS-CoV-2 (**e**) are shown. In **b** and **d**, significance was assessed by one-way analysis of variance (ANOVA) corrected for multiple comparisons using Tukey's method. The bars represent mean values  $\pm$  standard deviations. In **c** and **e**, longitudinal data are plotted over time continuously. Regression lines are shown as blue (previously infected) and purple (uninfected). The lines indicate cross-sectional averages from each group with shading representing 95% CI and are coloured accordingly. Significance was assessed using unpaired two-tailed *t*-test. TP 0:  $n = 37$ ; TP 1:  $n = 35$ ; TP 2:  $n = 30$ ; TP 3:  $n = 34$ ; TP 4:  $n = 31$ ; TP 5:  $n = 27$ ). Each dot represents a single individual. \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$  and \* $P < 0.05$ .

We found that mRNA vaccines induced high titres of virus-specific antibodies that declined over time, as previously reported<sup>6,11</sup> (Fig. [1b, c](#)). After the first vaccine dose,

over 97% of vaccinated participants developed virus-specific IgG titres, which increased to 100% after the second dose. IgG titres against the spike protein, spike subunit 1 and receptor-binding domain (RBD) peaked 7 days after the second dose of the vaccine (Fig. [1b,c](#)). No differences were observed in antibody levels between vaccinated participants of different sexes and after stratification by age (Extended Data Fig. [1a](#)). Consistent with previous reports<sup>[7,12](#)</sup>, we found that virus-specific IgG levels were significantly higher in the vaccinated group who were previously infected than in the vaccinated group who were uninfected (Extended Data Fig. [1b,c](#)). As expected, given the absence of sequences encoding nucleocapsid antigens in the mRNA vaccines, anti-nucleocapsid antibody titres remain stable over time for vaccinated individuals who were previously infected, and were not affected by vaccination in both the uninfected and the previously infected groups (Fig. [1b,c](#)). We next assessed plasma neutralization activity longitudinally against an authentic SARS-CoV-2 strain USA-WA1/2020 (lineage A), with a similar spike gene amino acid sequence as the Wuhan-Hu-1 sequence that was used for the mRNA vaccine design, by a 50% plaque-reduction neutralization (PRNT50) assay. Neutralization activity directly correlated with anti-spike and anti-RBD antibody titres, also peaking at 7 days after the second dose of the vaccine (Fig. [1d,e](#)). However, both groups displayed similar neutralization titres against the lineage A virus isolate at the peak of response (Fig. [1d,e](#)). Our data indicate that despite faster and more exuberant antibody responses to viral proteins by vaccinated individuals who were previously infected than vaccinated individuals who were uninfected, vaccination led to overall similar levels of NAbs after the second dose of the vaccine.

## Vaccine-induced T cell responses

A robust T cell response has also been linked to efficient protective immunity against SARS-CoV-2 (refs. [13,14,15](#)). Hence, we next longitudinally analysed spike-reactive and nucleocapsid-reactive T cell responses in vaccinated individuals. To detect low-frequency peptide-specific T cell populations, we first expanded antigen-specific T cells by stimulating PBMCs from vaccinated individuals with spike and nucleocapsid peptide pools ex vivo for 6 days, followed by restimulation with the same peptide pools and analysis of activation markers after 12 h. To cover the entire spike protein, two peptide pools (S-I and S-II) were used, while a single peptide pool was used for nucleocapsid stimulation. Spike-reactive CD4<sup>+</sup> and CD8<sup>+</sup> T cells increased over time following vaccination (Fig. [2a,b](#)), as evidenced by an increase in the number of cells expressing the activation markers CD38 and HLA-DR; no differences were observed between the previously infected and uninfected vaccinated groups. Consistent with previous reports<sup>[16,17](#)</sup>, spike-reactive CD4<sup>+</sup> T cell responses were comparable among full-length lineage A and P.1 virus isolates. By contrast, spike-reactive CD8<sup>+</sup> T cell responses were only observed to lineage A, and not P.1, 28 days after the second dose

of the vaccine, suggesting that spike-specific CD8<sup>+</sup> T cell responses can be affected by the mutations within the spike gene of SARS-CoV-2 variants (Fig. 2b). As expected, nucleocapsid-reactive T cells induced after stimulation with a nucleocapsid peptide pool derived from the lineage A virus isolate were primarily observed in vaccinated individuals who were previously infected (Fig. 2b). We also observed elevated numbers of nucleocapsid-reactive CD4<sup>+</sup> T cells in previously infected individuals at 28 days after the second dose of the vaccine, paralleling general activation of CD4<sup>+</sup> T cells (Fig. 2b, Extended Data Fig. 2a). Moreover, we observed increased counts of activated CD4<sup>+</sup> T cell, follicular helper T cell and antibody-secreting cells 28 days after the booster vaccine (Extended Data Fig. 2). Thus, T cell responses in vaccinated individuals display similar dynamics as antibody responses.

**Fig. 2: Temporal dynamics of anti-SARS-CoV-2 T cell immunity in vaccinated participants.**

 figure2

SARS-CoV-2 S-reactive CD4<sup>+</sup> and CD8<sup>+</sup> T cells after in vitro stimulation with SARS-CoV-2 S-I and S-II peptide pools and nucleoprotein peptide pool. **a**, Representative dot plots from four vaccinated individuals, 28 days after the second dose of the vaccine, showing the percentage of double-positive cells expressing HLA-DR and CD38 out of CD4<sup>+</sup> T cells (top) and CD8<sup>+</sup> T cells (bottom). Individuals who were previously infected to SARS-CoV-2 or uninfected are indicated by blue or purple shades, respectively. **b**, The percentage of double-positive cells, spike-reactive and nucleocapsid-reactive out of CD4<sup>+</sup> T cells (top) and CD8<sup>+</sup> T cells (bottom) over time post-vaccination. Individuals who were previously infected to SARS-CoV-2 or

uninfected are indicated by blue or purple dots, respectively. Each dot represents a single individual. Significance was assessed by one-way ANOVA corrected for multiple comparisons using Dunnett's method. Vaccination time points were compared with baseline. Stimulation values were subtracted from the respective non-stimulation condition. The bars represent the distribution of variables with quartiles and outliers. The horizontal bars indicate mean values. TP 0:  $n = 30$ ; TP 1:  $n = 34$ ; TP 2:  $n = 27$ ; TP 3:  $n = 27$ ; TP 4:  $n = 24$ ). Non-stim, non-stimulated PBMCs; nucleocapsid, PBMCs stimulated with the SARS-CoV-2 nucleocapsid protein pool derived from the ancestral lineage A virus, WA1, USA; spike, PBMCs stimulated with the SARS-CoV-2 spike protein pool derived from the ancestral strain lineage A, WA1, USA; spike (P.1), PBMCs stimulated with the SARS-CoV-2 spike protein pool derived from the P.1 variant. \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$  and \*\* $P < 0.01$ .

## Vaccine-induced NAb against variants

To investigate potential differences in NAb escape between the SARS-CoV-2 variants, we analysed the neutralization capacity of plasma samples from vaccinated individuals against a panel of 18 genetically distinct and authentic SARS-CoV-2 isolates. Among the isolates, 16 were from our Connecticut SARS-CoV-2 genomic surveillance programme representing variants from the same geographical region as our HCW cohort<sup>18</sup>. Our variant panel includes representatives of all lineages currently classified as variants of concern (B.1.1.7, B.1.351, P.1 and B.1.617.2) as well as lineages classified as variants of interest (B.1.427, B.1.429, B.1.525, B.1.526 and B.1.617.1)<sup>19</sup>. In addition, we selected lineages with key spike gene mutations (B.1.517 with N501T, and B.1 and R.1 with E484K)<sup>20</sup>, and included lineage A as a comparison (Fig. 3a). To help deconvolute the effects of individual mutations, we included four different B.1.526 isolates (labelled as B.1.526<sup>a-d</sup>) that represent different phylogenetic clades and key spike gene mutations (L452R, S477N and E484K) (Extended Data Fig. 3b), two different B.1.1.7 isolates with (B.1.1.7<sup>b</sup>) and without (B.1.1.7<sup>a</sup>; most common) E484K, and two B.1.351 isolates with (B.1.351<sup>b</sup>) and without (B.1.351<sup>a</sup>; most common) L18F (Fig. 3a, Extended Data Table 2). Except lineage A, all isolates (lineages B, P and R) have the spike gene D614G mutation located in the receptor-binding motif, which has been reported to have a modest effect on vaccine-elicited neutralization<sup>21</sup>. For each isolate, we highlight key spike protein amino acid differences in the antigenic regions of the amino-terminal domain (NTD), the RBD–ACE2 interface and the furin cleavage site (Fig. 3a). A full list of amino acid substitutions and deletions from all genes is provided in Extended Data Table 2. We used a PRNT50 assay to determine the neutralization titres of plasma collected 28 days after the second dose of the vaccine.

**Fig. 3: Effect of SARS-CoV-2 variants of concern on neutralization capacity of vaccinated participants.**

 figure3

**a**, Plasma neutralization titres against ancestral lineage A virus (WA1, USA) and locally circulating variants. Sixteen SARS-CoV-2 variants were isolated from nasopharyngeal swabs of individuals who were infected and an additional B.1.351 isolate was obtained from BEI. Neutralization capacity was assessed using plasma samples from vaccinated participants, 28 days after the second dose of the SARS-CoV-2 vaccine at the experimental sixfold serial dilutions (from 1:3 to 1:2,430). Key spike mutations within the distinct lineages and plasma neutralization titres (PRNT50) are also shown. Spike mutations are arranged across columns and the rows represent lineages. Significance was assessed by one-way ANOVA corrected for multiple

comparisons using Dunnett's method. Neutralization capacity of variants was compared with the neutralization capacity against the ancestral strain. The bars represent mean values  $\pm$  standard deviations, and the dotted line indicates the mean value of PRNT50 to the ancestral strain. Superscripted letters indicate different isolates belonging to the same lineage, with distinct spike amino acid changes.  $n = 32$ . FCS, furin cleavage site; RBM, receptor-binding motif. **b**, Estimated effect of individual mutations on plasma neutralization titres. Neutralization estimates (log PRNT50) and significance were tested with a linear mixed model with subject-level random effects. The dots represent the ratio of linear mixed model coefficients plus intercept to the intercept alone, the error bars represent the standard error, and the dotted line indicates no estimated effect on neutralization. **c**, Individual trajectories of plasma neutralization titres (PRNT50). Each line represents a single individual.  $n = 32$ . Variants were grouped giving specific spike mutations and are coloured accordingly. Superscripted letters indicate different isolates belonging to the same lineage, with distinct spike amino acid changes. \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$  and \* $P < 0.05$ .

When comparing vaccine-induced neutralization against the different isolates in comparison to the lineage A virus isolate, we observed significantly reduced PRNT50 titres for 12 out of the 17 isolates, and the rank order of reduced neutralization mostly clustered by key spike gene amino acid differences (Fig. 3a). Virus isolates with both the E484K and the N501Y (or N501T) mutations (B.1.351<sup>b</sup>, B.1.351<sup>a</sup>, B.1.1.7<sup>b</sup>, B.1 and P.1) reduced neutralization the most (4.6–6.0-fold decrease in PRNT50 titres). Virus isolates with the L452R mutation (B.1.617.1, B.1.429, B.1.526<sup>b</sup>, B.1.427 and B.1.617.2) were in the next grouping of decreased neutralization (2.5–4.1-fold decrease), which partially overlapped with isolates with E484K but without N501Y/T (B.1.525, R.1 and B.1.526<sup>c</sup>; 2.0–3.8-fold decrease). To further assess the effect of individual mutations, we constructed a linear mixed model with subject-level random effects to account for the differences in neutralization outcome (log-transformed PRNT50 titres) by each individual mutation as compared with lineage A (with no mutation) (Fig. 3b). From our model, we estimated that 8 of the 11 key spike gene mutations that we investigated had significant negative effects on neutralization, and that L452R (2.8-fold decrease in PRNT50 titres;  $P < 2 \times 10^{-16}$ ) and E484K/Q (2.0-fold decrease;  $P < 2 \times 10^{-16}$ ) had the greatest individual effects. As combinations of mutations can alter effects differently than the added value of each individually (that is, epistatic interactions), we also created a second linear mixed model that controlled for all of the individual mutations in the first model as well as three common combinations of key spike gene mutations found in our isolates: ΔH69/V70 and E484K, L452R and P681R, and E484K and N501Y. These combinations of mutations allowed us to assess whether the contribution of mutations together is synergistic, antagonistic or neither. Our model suggests that the ΔH69/V70 and E484K combination was synergistic (that is, decreased neutralization more than the added

effects of each;  $\beta = -0.182$ ;  $P = 0.005$ ), L452R and P681R was antagonistic (that is, decreased neutralization less than the added effects of each;  $\beta = 0.228$ ;  $P = 0.003$ ), and E484K and N501Y was neither (that is, neutralization was probably the sum of the individual effects of each;  $\beta = 0.060$ ;  $P = 0.248$ ). Thus, from our large panel of virus isolates, we find that virus genotype has an important role in vaccine-induced neutralization, with L452R probably having the largest individual effect, but the added effects of E484K and N501Y make viruses with this combination perhaps the most concerning for vaccines.

Although virus-specific factors may have a significant role in neutralization, differences in neutralization activity between individual vaccinated HCWs were much larger (up to approximately 2 log PRNT50 titres) than differences among virus isolates (mostly less than 1 log) (Fig. 3a). By tracking PRNT50 titres from each HCW, we found that the vaccinated individuals with high neutralization activity for the lineage A virus isolate are typically on the higher end of neutralization for all variants (Fig. 3c). Moreover, two vaccinated HCWs did not develop NAbs against any of the virus isolates, including lineage A (Fig. 3c), despite the production of virus-specific antibodies (Fig. 1a, b).

## Effect of previous infection on NAbs

To further understand the underlying factors that determine the levels of neutralization activity, we separated individuals by their previous SARS-CoV-2 infection status (that is, previously infected versus uninfected) and determined their neutralization titres to our panel of SARS-CoV-2 isolates. Previous infection occurred between April and December 2020, and for 3 out of 15 HCWs, we were able to identify the lineage of the virus, which were all B.1.3. While the rank order of virus isolates affecting neutralization remained mostly the same, we found that the plasma from vaccinated individuals who were previously infected generally had higher PRNT50 titres against the panel of SARS-CoV-2 isolates than vaccinated individuals who were uninfected (Fig. 4a, b). With the exception of virus isolates from lineages A, B.1.526<sup>a-c</sup> and R.1, which affected neutralization the least, all other assayed isolates had a significantly higher NAb response in vaccinated individuals who were previously infected (Fig. 4b); only virus isolates with the E484K and N501Y/T mutations still significantly reduced neutralization (Fig. 4a). For example, the lineage B.1.351<sup>b</sup> isolate (E484K and N501Y) decreased neutralization titres by 13.2-fold (compared with lineage A) in vaccinated individuals who were uninfected and by 3.7-fold in vaccinated individuals who were previously infected, whereas B.1.617.2 (L452R) went from 6.9-fold to 1.5-fold, and B.1.1.7<sup>a</sup> (N501Y) went from 3.4-fold to 0.8-fold decrease in individuals who were uninfected to individuals who were infected, respectively (Fig. 4a). Thus, our data suggest that plasma neutralization activity against SARS-CoV-2 variants is improved in vaccinated individuals who were previously infected with the virus.

**Fig. 4: Comparison of neutralizing activity in vaccinated HCWs who were previously infected or not to SARS-CoV-2.**

figure4

Plasma neutralization titres against ancestral lineage A virus (WA1, USA) and locally circulating variants of concern or interest, and other lineages. Sixteen SARS-CoV-2 variants were isolated from nasopharyngeal swabs of individuals who were infected and an additional B.1.351 isolate was obtained from BEI. Neutralization capacity between vaccinated participants who were previously infected or not with SARS-CoV-2 is also shown, and was accessed using plasma samples from vaccinated participants, 28 days after the second dose of the SARS-CoV-2 vaccine at the experimental sixfold serial dilutions (from 1:3 to 1:2,430). **a**, Neutralization titre among vaccinated

individuals. Significance was assessed by one-way ANOVA corrected for multiple comparisons using Dunnett's method. Neutralization capacity to the variants was compared with the neutralization capacity against the ancestral strain. The bars represent mean values  $\pm$  standard deviations; the dotted line indicates the mean value of PRNT50 to the ancestral strain. Variants were grouped giving specific spike mutations and are coloured accordingly. **b**, Comparison of the neutralization titre among vaccinated participants who were previously infected or not to SARS-CoV-2. Significance was accessed using unpaired two-tailed *t*-test. The bars represent mean values  $\pm$  standard deviations. (–) Vaccinated, uninfected:  $n = 17$ ; (+) Vaccinated, previously infected:  $n = 15$ . Each dot represents a single individual. \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$  \*\* $P < 0.01$  and \* $P < 0.05$ .

## Discussion

Human NAbs against SARS-CoV-2 can be categorized as belonging to four classes on the basis of their target regions on the RBD. Although the RBD is immunodominant, there is evidence for a substantial role of other spike regions in antigenicity, most notably the NTD supersite<sup>22,23,24</sup>. These antibodies target epitopes that are closely associated with NTD and RBD residues L18 and ΔH69/V70, and K417, L452, S477, T478, E484 and N501. Previous studies using pseudovirus constructs reported a significant effect of single spike gene amino acid substitutions, including S477N and E484K located at the RBD–ACE2 interface, in the neutralization activity of plasma from vaccinated individuals<sup>1,2,3,4,5,25</sup>.

Using a large panel of genetically diverse authentic SARS-CoV-2 isolates, we found that lineages with E484K and N501Y/T led to the most severe decreases in mRNA vaccine-induced neutralization (more than tenfold in vaccinated individuals who were previously uninfected). This group includes B.1.351 (Beta) and P.1 (Gamma), further supporting their importance in regards to vaccines. We also found that a generic lineage B.1 isolate with E484K and N501T, and a rare B.1.1.7 (Alpha) isolate with E484K (also with the common N501Y mutation) have similar effects on neutralization as B.1.351 and P.1. While the combinations of mutations in the B.1 and B.1.1.7 with E484K isolates probably do not increase transmissibility, the additive effects of these two mutations support that surveillance programmes should track all viruses with E484K and N501Y/T mutations in addition to variants of concern and/or interest.

While we estimate that the L452R mutation has the greatest individual effect on neutralization, lineages with this mutation, including B.1.617.2 (Delta), are less concerning for NAb escape than the E484K–N501Y/T combination. From our data, we expect that most fully vaccinated individuals will be protected against B.1.617.2, and that the rise in vaccine-breakthroughs associated with this variant are more probably associated with its high transmissibility<sup>26,27,28,29</sup>.

The discrepancies of our results compared with other studies, including ref. <sup>6</sup>, may point to the importance of using fully intact, authentic virus for neutralization assays to detect the effects of epistasis among virus mutations on neutralization assays.

Nevertheless, it remains possible that additional factors also contribute to some of the discrepancies between our observations and those of previous studies, including the presence of additional mutations in the membrane and envelope, as well as the composition of our cohorts, predominantly young white women. Differences between cohorts could also account for subtle differences in T cell responses observed in our study versus the one recently reported by ref. <sup>16</sup> and ref. <sup>17</sup>. While we observed decreased cross-reactivity of spike-reactive CD8<sup>+</sup> T cells against P.1 spike peptides, the above studies found that T cell responses are largely preserved against variants<sup>16,17</sup>. In addition to cohort composition, we used overlapping peptide pools in our assays, and it remains possible that our T cell assay failed to detect under-represented T cell clones affected by variant sequences when sampled in the presence of the majority of conserved peptides. Furthermore, ref. <sup>16</sup> used an AdV1.26 vaccine, whereas in this study and ref. <sup>17</sup>, mRNA vaccines were used. Overall, our data point to a necessity of active monitoring of T cell reactivity in the context of SARS-CoV-2 evolution.

The magnitude of the antibody titres in patients with COVID-19 following natural infection has been directly correlated with length of infection and severity<sup>30</sup>. Here we found that vaccinated individuals who were previously infected display an increased resilience in antibody responses against both ‘single’ and a combination of substitutions in the RBD region, which otherwise severely decreased neutralization activity of vaccinated individuals who were uninfected. Our observations of the effect of pre-existing immunity in vaccinated individuals on their ability to neutralize variants could be explained by the time window between the initial exposure (infection) and vaccination. Moreover, on the basis of timing of previous infection (before the emergence of tested variants) and confirmation by sequencing (three previous infections with B.1.3), we believe that the virus lineage of the infection probably did not have a major effect on our findings. Our observations provide an important rationale for worldwide efforts in characterizing the contribution of pre-existing SARS-CoV-2 immunity to the outcome of various vaccination strategies. Along with recently introduced serological tests<sup>31</sup>, such studies could inform evidence-based risk evaluation, patient monitoring, adaptation of containment methods, and vaccine development and deployment. Finally, these findings suggest that an additional third dose of a vaccine may be beneficial to confer higher protection against SARS-CoV-2 lineages such as B.1.351 and P.1.

## Methods

## **Ethics statement**

This study was approved by the Yale Human Research Protection Program Institutional Review Board (IRB protocol ID 2000028924). Informed consent was obtained from all enrolled vaccinated HCWs. The IRB from the Yale University Human Research Protection Program determined that the RT-qPCR testing and sequencing of de-identified remnant COVID-19 clinical samples conducted in this study were not research involving human participants (IRB protocol ID 2000028599).

## **HCW volunteers**

Forty HCW volunteers from the Yale-New Haven Hospital were enrolled and included in this study. The volunteers received the mRNA vaccine (Moderna or Pfizer) between November 2020 and January 2021. Vaccinated donors were stratified in two major groups, previously infected with SARS-CoV-2 (recovered) or uninfected (naive), confirmed by RT-qPCR (10 April 2020 to 31 December 2020) and serology (2 April 2020 to 11 March 2020). None of the participants experienced serious adverse effects after vaccination. HCWs were followed serially post-vaccination. Plasma and PBMC samples were collected at baseline (previous to vaccination), 7 and 28 after the first vaccination dose, and 7, 28 and 70 days after the second vaccination dose.

Demographic information was aggregated through a systematic review of the electronic health record (EHR) and was used to construct Extended Data Table 1. The clinical data were collected using EPIC EHR May 2020 and REDCap 9.3.6 software. Blood acquisition was performed and recorded by a separate team. Clinical information and time points of collection information of vaccinated HCWs were not available until after processing and analysing raw data by flow cytometry and ELISA. ELISA, neutralizations and flow cytometry analyses were blinded.

## **Isolation of plasma and PBMCs**

Whole blood was collected in heparinized CPT blood vacutainers (BDAM362780, BD) and kept on gentle agitation until processing. All blood was processed on the day of collection in a single step standardized method. Plasma samples were collected after centrifugation of whole blood at 600g for 20 min at room temperature without a break. The undiluted plasma was transferred to 15-ml polypropylene conical tubes, and aliquoted and stored at -80 °C for subsequent analysis. The PBMC layer was isolated according to the manufacturer's instructions. Cells were washed twice with PBS before counting. Pelleted cells were briefly treated with ACK lysis buffer for 2 min and then counted. Percentage viability was estimated using standard trypan blue staining and an automated cell counter (AMQAX1000, Thermo Fisher). PBMCs were stored at -80 °C for subsequent analysis.

## SARS-CoV-2-specific antibody measurements

ELISAs were performed as previously described<sup>32</sup>. In short, Triton X-100 and RNase A were added to serum samples at final concentrations of 0.5% and 0.5 mg/ml, respectively, and incubated at room temperature for 30 min before use, to reduce risk from any potential virus in serum. MaxiSorp plates (96 wells; 442404, Thermo Scientific) were coated with 50 µl per well of recombinant SARS-CoV-2 STotal (SPN-C52H9-100 µg, ACROBiosystems), S1 (S1N-C52H3-100 µg, ACROBiosystems), RBD (SPD-C52H3-100 µg, ACROBiosystems) and the nucleocapsid protein (NUN-C5227-100 µg, ACROBiosystems) at a concentration of 2 µg/ml in PBS and were incubated overnight at 4 °C. The coating buffer was removed, and plates were incubated for 1 h at room temperature with 200 µl of blocking solution (PBS with 0.1% Tween-20 and 3% milk powder). Plasma was diluted serially at 1:100, 1:200, 1:400 and 1:800 in dilution solution (PBS with 0.1% Tween-20 and 1% milk powder), and 100 µl of diluted serum was added for 2 h at room temperature. Human anti-spike (SARS-CoV-2 human anti-spike (AM006415) (91351, Active Motif) and anti-nucleocapsid SARS-CoV-2 human anti-nucleocapsid (1A6) (MA5-35941, Active Motif) were serially diluted to generate a standard curve. Plates were washed three times with PBS-T (PBS with 0.1% Tween-20) and 50 µl of HRP anti-human IgG antibody (1:5,000; A00166, GenScript) diluted in dilution solution added to each well. After 1 h of incubation at room temperature, plates were washed six times with PBS-T. Plates were developed with 100 µl of TMB Substrate Reagent Set (555214, BD Biosciences) and the reaction was stopped after 5 min by the addition of 2 N sulfuric acid. Plates were then read at a wavelength of 450 nm and 570 nm.

## T cell stimulation

For the in vitro stimulation, PBMCs were stimulated with HLA class I and HLA-DR peptide pools at the concentration of 1–10 µg ml<sup>-1</sup> per peptide and cultured for 7 days. On day 0, PBMCs were thawed, counted and plated in a total of 5–8 × 10<sup>5</sup> cells per well in 200 µl of RPMI 1640 medium (Gibco) supplemented with 1% sodium pyruvate (NEAA), 100 U/ml penicillin–streptomycin (Biochrom) and 10% FBS at 37 °C and 5% CO<sub>2</sub>. On day 1, cells were washed and the stimulation was performed with:

PepMix SARS-CoV-2 spike glycoprotein pool 1 and pool 2 (GenScript), PepMix P.1 SARS-CoV-2 spike glycoprotein pool 1 and pool 2 (JPT) and PepMix SARS-CoV-2 nucleocapsid protein (JPT). Stimulation controls were performed with PBS (unstimulated). Peptide pools were used at 1 µg ml<sup>-1</sup> per peptide. Incubation was performed at 37 °C, 5% CO<sub>2</sub> for 6 days. On day 6, cells were restimulated with 10 µg ml<sup>-1</sup> per peptide and subsequently incubated for 12 h, with the last 6 h being in the presence of 10 µg ml<sup>-1</sup> brefeldin A (Sigma-Aldrich). Following this incubation, cells were washed with PBS 2 mM EDTA and prepared for analysis by flow cytometry.

## Flow cytometry

Antibody clones and vendors were as follows: BB515 anti-hHLA-DR (G46-6, 1:400; BD Biosciences), BV605 anti-hCD3 (UCHT1, 1:300; BioLegend), BV785 anti-hCD19 (SJ25C1, 1:300; BD Biosciences), BV785 anti-hCD4 (SK3, 1:200; BioLegend), APCFire750 or BV711 anti-hCD8 (SK1, 1:200; BioLegend), AlexaFluor 700 anti-hCD45RA (HI100, 1:200; BD Biosciences), PE anti-hPD1 (EH12.2H7, 1:200; BioLegend), APC or PE-CF594 anti-hTIM3 (F38-2E2, 1:50; BioLegend), BV711 anti-hCD38 (HIT2, 1:200; BioLegend), BB700 anti-hCXCR5 (RF8B2, 1:50; BD Biosciences), PE-CF594 anti-hCD25 (BC96, 1:200; BD Biosciences), AlexaFluor 700 anti-hTNFa (MAb11, 1:100; BioLegend), PE or APC/Fire750 anti-hIFN $\gamma$  (4S.B3, 1:60; BioLegend), FITC anti-hGranzymeB (GB11, 1:200; BioLegend), BV785 anti-hCD19 (SJ25C1, 1:300; BioLegend), BV421 anti-hCD138 (MI15, 1:300; BioLegend), AlexaFluor700 anti-hCD20 (2H7, 1:200; BioLegend), AlexaFluor 647 anti-hCD27 (M-T271, 1:350; BioLegend), PE/Dazzle594 anti-hIgD (IA6-2, 1:400; BioLegend), Percp/Cy5.5 anti-hCD137 (4B4-1, 1:150; BioLegend) and PE anti-CD69 (FN-50, 1:200; BioLegend), and APC anti-hCD40L (24-31, 1:100; BioLegend). In brief, freshly isolated PBMCs were plated at  $1\text{--}2 \times 10^6$  cells per well in a 96-well U-bottom plate. Cells were resuspended in Live/Dead Fixable Aqua (Thermo Fisher) for 20 min at 4 °C. Following a wash, cells were blocked with Human TruStan FcX (BioLegend) for 10 min at room temperature. Cocktails of desired staining antibodies were added directly to this mixture for 30 min at room temperature. For secondary stains, cells were first washed and supernatant aspirated; then to each cell pellet, a cocktail of secondary markers was added for 30 min at 4 °C. Before analysis, cells were washed and resuspended in 100  $\mu\text{l}$  4% PFA for 30 min at 4 °C. Following this incubation, cells were washed and prepared for analysis on an Attune NXT (Thermo Fisher). Data were analysed using FlowJo software version 10.6 software (Tree Star). The specific sets of markers used to identify each subset of cells are summarized in Extended Data Fig. 4.

## Cell lines and virus

TMPRSS2-VeroE6 kidney epithelial cells were cultured in DMEM supplemented with 1% sodium pyruvate (NEAA) and 10% FBS at 37 °C and 5% CO<sub>2</sub>. The cell line was obtained from the American Type Culture Collection and tested negative for contamination with mycoplasma. SARS-CoV-2 lineage A (USA-WA1/2020) was obtained from BEI Resources (no. NR-52281) and was amplified in TMPRSS2-VeroE6. Cells were infected at a multiplicity of infection of 0.01 for 3 days to generate a working stock, and after incubation, the supernatant was clarified by centrifugation (450g for 5 min), and filtered through a 0.45- $\mu\text{m}$  filter. The pelleted virus was then resuspended in PBS and aliquoted for storage at -80 °C. Viral titres were measured by standard plaque assay using TMPRSS2-VeroE6. In brief, 300  $\mu\text{l}$  of serial fold virus dilutions were used to infect Vero E6 cells in MEM supplemented

with NaHCO<sub>3</sub>, 4% FBS and 0.6% Avicel RC-581. Plaques were resolved at 48 h post-infection by fixing in 10% formaldehyde for 1 h followed by 0.5% crystal violet in 20% ethanol staining. Plates were rinsed in water to plaques enumeration. All experiments were performed in a biosafety level 3 laboratory with approval from the Yale Environmental Health and Safety office.

## SARS-CoV-2 variant sequencing and isolation

SARS-CoV-2 samples were sequenced as part of the Yale SARS-CoV-2 Genomic Surveillance Initiative's weekly surveillance programme in Connecticut, USA<sup>33</sup>. Lineages were sequenced and isolated as previously described<sup>34</sup>. In brief, nucleic acid was extracted from de-identified remnant nasopharyngeal swabs and tested with our multiplexed RT–qPCR variant assay to select samples with a N1 cycle threshold value of 35 or lower for sequencing<sup>35,36</sup>. Libraries were prepared with a slightly adjusted version of the Illumina COVIDSeq Test RUO version. The Yale Center for Genome Analysis sequenced pooled libraries of up to 96 samples on the Illumina NovaSeq (paired-end 150). Data were analysed and consensus genomes were generated using iVar (version 1.3.1)<sup>37</sup>. Variants of interest and concern, lineages with mutations of concern (E484K), as well as other lineages as controls were selected for virus isolation. In total, 16 viruses were isolated belonging to 12 lineages (Extended Data Fig. 3, Extended Data Table 2). In addition, ancestral lineage A virus and lineage B.1.351 virus were obtained from BEI.

Samples selected for virus isolation were diluted 1:10 in DMEM and then filtered through a 45-μm filter. The samples were tenfold serially diluted from 1:50 to 1:19,531,250. The dilution was subsequently incubated with TMPRSS2-Vero E6 in a 96-well plate and adsorbed for 1 h at 37 °C. After adsorption, replacement medium was added, and cells were incubated at 37 °C for up to 5 days. Supernatants from cell cultures with cytopathic effect were collected, frozen, thawed and subjected to RT–qPCR. Fresh cultures were inoculated with the lysates as described above for viral expansion. Viral infection was subsequently confirmed through reduction of cycle threshold values in the cell cultures with the multiplex variant qPCR assay. Expanded viruses were resequenced following the same method as described above and genome sequences were uploaded to GenBank (Supplementary Data Table 2), and the aligned consensus genomes are available on GitHub ([https://github.com/grubaughlab/paper\\_2021\\_Nab-variants](https://github.com/grubaughlab/paper_2021_Nab-variants)). Nextclade v1.5.0 (<https://clades.nextstrain.org/>) was used to generate a phylogenetic tree (Extended Data Fig. 3) and to compile a list of amino acid changes in the virus isolates as compared with the Wuhan-Hu-1 reference strain (Extended Data Table 2). Key spike amino acid differences were identified based on the outbreak.info mutation tracker<sup>20</sup>.

## Neutralization assay

Sera from vaccinated HCWs were isolated as described before and then heat treated for 30 min at 56 °C. Sixfold serially diluted plasma, from 1:3 to 1:2,430 were incubated with SARS-CoV-2 variants, for 1 h at 37 °C. The mixture was subsequently incubated with TMPRSS2-VeroE6 in a 12-well plate for 1 h, for adsorption. Then, cells were overlayed with MEM supplemented with NaHCO<sub>3</sub>, 4% FBS and 0.6% Avicel mixture. Plaques were resolved at 40 h post-infection by fixing in 10% formaldehyde for 1 h followed by staining in 0.5% crystal violet. All experiments were performed in parallel with sera from baseline controls, in an established viral concentration to generate 60–120 plaques per well.

## Statistical analysis

All analyses of patient samples were conducted using GraphPad Prism 8.4.3, JMP 15 and R 3.4.3. Multiple group comparisons were analysed by running parametric (ANOVA) statistical tests. Multiple comparisons were corrected using Tukey's and Dunnett's tests as indicated in the figure legends. For the comparison between stable groups, two-sided unpaired *t*-test was used for the comparison. The effect of spike mutations was assessed using a linear mixed model with an outcome of log-transformed PRNT50 and random effects accounting for each individual participant. This was done using the ‘lme4’ package in R 4.0.1 (ref. [38](#)).

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

The data generated during the current study are available as indicated. All of the background information for HCW participants and data generated in this study are included in the Supplementary Data file. All of the genome information for SARS-CoV-2 variants used in this study are available in Extended Data Table [2](#), and the aligned consensus genomes are available on GitHub ([https://github.com/grubaughlab/paper\\_2021\\_Nab-variants](https://github.com/grubaughlab/paper_2021_Nab-variants)). Additional correspondence and requests for materials should be addressed to the corresponding author (A.I.).

## References

1. 1.

Cele, S. et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **593**, 142–146 (2021).

2. 2.

Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383.e9 (2021).

3. 3.

Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).

4. 4.

Chen, R. E. et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).

5. 5.

Wang, P. et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).

6. 6.

Stamatatos, L. et al. mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021).

7. 7.

Jackson, L. A. et al. An mRNA vaccine against SARS-CoV-2—preliminary report. *N. Engl. J. Med.* **383**, 1920–1931 (2020).

8. 8.

Walsh, E. E. et al. Safety and immunogenicity of two RNA-based Covid-19 vaccine candidates. *N. Engl. J. Med.* **383**, 2439–2450 (2020).

9. 9.

Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).

10. 10.

Sabino, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).

11. 11.

Jalkanen, P. et al. COVID-19 mRNA vaccine induced antibody responses against three SARS-CoV-2 variants. *Nat. Commun.* **12**, 3991 (2021).

12. 12.

Sahin, U. et al. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. *Nature* **586**, 594–599 (2020).

13. 13.

Braun, J. et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020).

14. 14.

Dan, J. M. et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021).

15. 15.

Zhao, J. et al. Airway memory CD4<sup>+</sup> T cells mediate protective immunity against emerging respiratory coronaviruses. *Immunity* **44**, 1379–1391 (2016).

16. 16.

Alter, G. et al. Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans. *Nature* **596**, 268–272 (2021).

17. 17.

Tarke, A. et al. Impact of SARS-CoV-2 variants on the total CD4<sup>+</sup> and CD8<sup>+</sup> T cell reactivity in infected or vaccinated individuals. *Cell Rep. Med.* **2**, 100355 (2021).

18. 18. Connecticut SARS-CoV-2 variant surveillance. *COVID Tracker* <https://covidtrackerct.com/> (2020).
19. 19. CDC. SARS-CoV-2 variant classifications and definitions. *CDC* <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html> (2021).
20. 20. SARS-CoV-2 (hCoV-19) mutation reports. *outbreak.info* <https://outbreak.info/situation-reports> (2021).
21. 21. Zou, J. et al. The effect of SARS-CoV-2 D614G mutation on BNT162b2 vaccine-elicited neutralization. *NPJ Vaccines* **6**, 44 (2021).
22. 22. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).
23. 23. Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).
24. 24. McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347 (2021).
25. 25. Xie, X. et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **27**, 620–621 (2021).
26. 26.

Bernal, J. L. et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594 (2021).

27. 27.

Bolze, A. et al. Rapid displacement of SARS-CoV-2 variant B.1.1.7 by B.1.617.2 and P.1 in the United States. Preprint at <https://doi.org/10.1101/2021.06.20.21259195> (2021).

28. 28.

Lustig, Y. et al. Neutralising capacity against Delta (B.1.617.2) and other variants of concern following Comirnaty (BNT162b2, BioNTech/Pfizer) vaccination in health care workers, Israel. *Euro Surveill.* **26**, 2100557 (2021).

29. 29.

Challen, R. et al. Early epidemiological signatures of novel SARS-CoV-2 variants: establishment of B.1.617.2 in England. Preprint at <https://doi.org/10.1101/2021.06.05.21258365> (2021).

30. 30.

Lucas, C. et al. Delayed production of neutralizing antibodies correlates with fatal COVID-19. *Nat. Med.* **27**, 1178–1186 (2021).

31. 31.

US Food & Drug Administration. EUA authorized serology test performance. *FDA* <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance> (2021).

32. 32.

Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* **26**, 1033–1036 (2020).

33. 33.

Kalinich, C. C. et al. Real-time public health communication of local SARS-CoV-2 genomic epidemiology. *PLoS Biol.* **18**, e3000869 (2020).

34. 34.

Mao, T. et al. A stem-loop RNA RIG-I agonist confers prophylactic and therapeutic protection against acute and chronic SARS-CoV-2 infection in mice. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.06.16.448754> (2021).

35. 35.

Vogels, C. B. F. et al. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol.* **19**, e3001236 (2021).

36. 36.

Vogels, C. B. F., Fauver, J. R. & Grubaugh, N. D. Multiplexed RT-qPCR to screen for SARS-CoV-2 B.1.1.7, B.1.351, and P.1 variants of concern V.3. *Protocols* <https://doi.org/10.17504/protocols.io.br9vm966> (2021).

37. 37.

Grubaugh, N. D. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).

38. 38.

Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

## Acknowledgements

We thank M. Linehan for technical and logistical assistance, D. Mucida for discussions, and M. Suchard and W. Hanage for statistics advice. This work was supported by the Women’s Health Research at Yale Pilot Project Program (A.I.), Fast Grant from Emergent Ventures at the Mercatus Center (A.I. and N.D.G.), Mathers Foundation, the Ludwig Family Foundation, the Department of Internal Medicine at the Yale School of Medicine, Yale School of Public Health and the Beatrice Kleinberg Neuwirth Fund. A.I. is an investigator of the Howard Hughes Medical Institute. C.L. is a Pew Latin American Fellow. C.B.F.V. is supported by NWO Rubicon 019.181EN.004.

## Author information

Author notes

1. These authors contributed equally: Carolina Lucas, Chantal B. F. Vogels, Inci Yildirim
2. These authors jointly supervised this work: Saad B. Omer, Nathan D. Grubaugh, Akiko Iwasaki

## Affiliations

1. Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA

Carolina Lucas, Peiwen Lu, Valter Monteiro, Jeff R. Gehlhausen, Julio Silva, Alexandra Tabachnikova, Mario A. Peña-Hernandez, Subhasis Mohanty, Jiefang Huang, Albert C. Shaw & Akiko Iwasaki

2. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA

Chantal B. F. Vogels, Jessica E. Rothman, M. Catherine Muenker, Mallory I. Breban, Joseph R. Fauver, Mary Petrone, Isabel Ott, Anne Watkins, Chaney Kalinich, Tara Alpert, Anderson Brito, Rebecca Earnest, Albert I. Ko, Saad B. Omer & Nathan D. Grubaugh

3. Department of Pediatric, Section of Infectious Diseases and Global Health, Yale University School of Medicine, New Haven, CT, USA

Inci Yildirim

4. Yale Institute for Global Health, Yale University, New Haven, CT, USA

Inci Yildirim & Saad B. Omer

5. Department of Dermatology, Yale University School of Medicine, New Haven, CT, USA

Jeff R. Gehlhausen

6. Department of Medicine, Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT, USA

Melissa Campbell, Subhasis Mohanty, Jiefang Huang, Albert C. Shaw, Albert I. Ko & Saad B. Omer

7. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

Nathan D. Grubaugh

8. Howard Hughes Medical Institute, Chevy Chase, MD, USA

Akiko Iwasaki

9. Connecticut State Department of Public Health, Rocky Hill, CT, USA

Claire Pearson, Anthony Muyombwe, Randy Downing & Jafar Razeq

10. Murphy Medical Associates, Greenwich, CT, USA

Steven Murphy, Caleb Neal, Eva Laszlo & Ahmad Altajar

11. Yale Center for Genome Analysis, Yale University, New Haven, CT, USA

Irina Tikhonova, Christopher Castaldi, Shrikant Mane & Kaya Bilguvar

12. Department of Laboratory Medicine, Yale New Haven Hospital, New Haven, CT, USA

Nicholas Kerantzas

13. Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, CT, USA

David Ferguson & Wade Schulz

14. Department of Laboratory Medicine, Yale School of Medicine, New Haven, CT, USA

Wade Schulz

15. Departments of Laboratory Medicine and Medicine, Yale University School of Medicine, New Haven, CT, USA

Marie Landry & David Peaper

## **Consortia**

### **Yale SARS-CoV-2 Genomic Surveillance Initiative**

- Claire Pearson
- , Anthony Muyombwe
- , Randy Downing
- , Jafar Razeq
- , Mary Petrone
- , Isabel Ott
- , Anne Watkins
- , Chaney Kalinich
- , Tara Alpert
- , Anderson Brito
- , Rebecca Earnest
- , Steven Murphy
- , Caleb Neal
- , Eva Laszlo
- , Ahmad Altajar
- , Irina Tikhonova
- , Christopher Castaldi
- , Shrikant Mane
- , Kaya Bilguvar
- , Nicholas Kerantzas
- , David Ferguson
- , Wade Schulz
- , Marie Landry
- & David Peaper

## Contributions

C.L., C.B.F.V., I.Y., S.B.O., N.D.G. and A.I. conceived the study. C.L., J.S., A.T., M.A.P.-H., S.M. and J.H. collected and processed patient PBMC and plasma samples. C.L. and V.M. performed SARS-CoV-2-specific antibody ELISAs. C.L. isolated SARS-CoV-2 variants and performed the neutralization assays. C.L., P.L. and J.R.G. performed the flow cytometry and the flow data analyses. I.Y. and M.C. collected epidemiological and clinical data. C.B.F.V., M.I.B., J.R.F. and Y.S.G.S.N. performed virus sequencing. M.C.M. processed and stored patient specimens. M.C. assisted in the identification and enrolment of volunteers. C.L., C.B.F.V., J.E.R. and N.D.G. analysed the data. C.L., C.B.F.V., N.D.G. and A.I. drafted the manuscript. All authors reviewed and approved the manuscript. N.D.G. and A.I. secured funds and supervised the project.

## Corresponding author

Correspondence to [Akiko Iwasaki](#).

## Ethics declarations

### Competing interests

A.I. served as a consultant for Spring Discovery, Boehringer Ingelheim and Adaptive Biotechnologies. I.Y. reported being a member of the mRNA-1273 Study Group and has received funding to her institution to conduct clinical research from BioFire, MedImmune, Regeneron, PaxVax, Pfizer, GSK, Merck, Novavax, Sanofi-Pasteur and Micron. N.D.G. is a consultant for Tempus Labs to develop infectious disease diagnostic assays. All other authors declare no competing interests.

### Additional information

**Peer review information** *Nature* thanks Pei-Yong Shi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Extended data figures and tables

#### [Extended Data Fig. 1 Correlation of virus-specific antibodies with age and sex of participants.](#)

**a, b**, Plasma reactivity to S protein and RBD in vaccinated participants measured over time by ELISA. HCW participants received 2 doses of the mRNA vaccines and plasma samples were collected at the indicated time points (TP). Baseline, previously to vaccination; 1 Time point, 7 days post 1 dose; 2 Time point, 28 days post 1 dose; 3 Time point, 7 days post 2 dose; 4 Time point, 28 days post 2 dose; 5 Time point, 70 days post 2 dose. **a**, Anti-S (left) and Anti-RBD (right) IgG levels stratified by vaccinated participants accordingly to age and sex. Significance was assessed using unpaired two-tailed t-test. Boxes represent variables' distribution with quartiles and outliers. Horizontal bars, mean values. **b**, Anti-S, Anti-S1, Anti-RBD and Anti-N IgG comparison in vaccinated participants previously infected or not to SARS-CoV-2. Longitudinal data plotted over time. Significance was assessed using unpaired two-tailed t-test. Boxes represent mean values  $\pm$  standard deviations. TP, vaccination time point. Anti-S IgG (TP0, n=37; TP1, n=35; TP2, n=30; TP3, n=34; TP4, n=34; TP5, n=28). Anti-S1 IgG (TP0, n=37; TP1, n=35; TP2, n=30; TP3, n=34; TP4, n=34; TP5, n=27). Anti-RBD IgG (TP0, n=37; TP1, n=35; TP2, n=30; TP3, n=34; TP4, n=34);

TP5, n=27). Anti-N IgG (TP0, n=37; TP1, n=35; TP2, n=30; TP3, n=34; TP4, n=34; TP5, n=27). S, spike. S1, spike subunit 1. RBD, receptor binding domain. N, nucleocapsid. Each dot represents a single individual. \*\*\*p < .0001 \*\*\*p < .001 \*\*p < .01\*p < .05.

### Extended Data Fig. 2 Cellular immune profiling post SARS-CoV-2 vaccination.

**a–c**, Immune cell subsets of interest, plotted as a percentage of a parent population over time according to the vaccination time points. HCW participants received 2 doses of the mRNA vaccines and PBMCs samples were collected as at the indicated time points (TP). Baseline, previously to vaccination; 1 Time point, 7 days post 1 dose; 2 Time point, 28 days post 1 dose; 3 Time point, 7 days post 2 dose; 4 Time point, 28 days post 2 dose. Percentage of activated T cell subsets (**a**), B cell subsets (**b**) and Tfh cells (**c**) among vaccinated individuals over time. Individuals previously infected to SARS-CoV-2 or uninfected are indicated by blue or purple dots, respectively. Each dot represents a single individual. Significance was assessed by One-way ANOVA corrected for multiple comparisons using Dunnett's method. Vaccination time points were compared with baseline. Boxes represent variables' distribution with quartiles and outliers. Horizontal bars, mean values. TP, vaccination time point (TP0, n = 29; TP1, n=33; TP2, n =26; TP3, n = 13; TP4, n=25). \*\*\*p < .001 \*\*p < .01\*p < .05.

### Extended Data Fig. 3 Maximum likelihood phylogeny of SARS-CoV-2 genomes of cultured virus isolates.

**a**, Nextclade (<https://clades.nextstrain.org/>) was used to generate a phylogenetic tree to show evolutionary relations between the cultured virus isolates used in this study and other publicly available SARS-CoV-2 genomes. Branches are colored by Pango Lineage, and labelled according to the WHO naming scheme. Highlighted are the cultured virus isolates used in this study. **b**, Enlarged section of the phylogenetic tree highlighting spike amino acid changes in the B.1.526 (iota) lineage viruses belonging to different clades.

### Extended Data Fig. 4 Gating strategies.

Gating strategies are shown for the key cell populations described in Fig. 2 and Extended Data Fig. 2. **a**, Leukocyte gating strategy to identify lymphocytes. T cell surface staining gating strategy to identify CD4<sup>+</sup> and CD8<sup>+</sup> T cells, TCR-activated T cells and follicular T cells. **b**, B cell surface staining gating strategy to identify B cells subsets.

### **Extended Data Table 1 SARS-CoV-2 vaccinated cohort**

## **Extended Data Table 2 Amino acid changes identified in cultured SARS-CoV-2 isolates**

## **Supplementary information**

### **Supplementary Data 1**

Detailed clinical and immunological data for each patient. Clinical information, demographics and exact counts for immunological data.

### **Reporting Summary**

### **Peer Review File**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Lucas, C., Vogels, C.B.F., Yildirim, I. *et al.* Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature* **600**, 523–529 (2021).  
<https://doi.org/10.1038/s41586-021-04085-y>

- Received: 10 July 2021
- Accepted: 30 September 2021
- Published: 11 October 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04085-y>

### **Share this article**

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Further reading

- [\*\*The germinal centre B cell response to SARS-CoV-2\*\*](#)

- Brian J. Laidlaw
- Ali H. Ellebedy

*Nature Reviews Immunology* (2021)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04085-y>.

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 20 October 2021](#)

# Hybrid immunity improves B cells and antibodies against SARS-CoV-2 variants

- [Emanuele Andreano](#) [ORCID: orcid.org/0000-0003-0088-5788<sup>1</sup>](#),
- [Ida Paciello<sup>1</sup>](#),
- [Giulia Piccini](#) [ORCID: orcid.org/0000-0001-8090-7910<sup>2</sup>](#),
- [Noemi Manganaro](#) [ORCID: orcid.org/0000-0001-8512-4704<sup>1</sup>](#),
- [Piero Pileri](#) [ORCID: orcid.org/0000-0002-2888-8748<sup>1</sup>](#),
- [Inesa Hysen<sup>1,3</sup>](#),
- [Margherita Leonardi<sup>2,3</sup>](#),
- [Elisa Pantano<sup>1</sup>](#),
- [Valentina Abbiento<sup>1</sup>](#),
- [Linda Benincasa<sup>3</sup>](#),
- [Ginevra Giglioli<sup>3</sup>](#),
- [Concetta De Santi<sup>1</sup>](#),
- [Massimiliano Fabbiani<sup>4</sup>](#),
- [Ilaria Rancan<sup>4,5</sup>](#),
- [Mario Tumbarello<sup>4,5</sup>](#),
- [Francesca Montagnani<sup>4,5</sup>](#),
- [Claudia Sala<sup>1</sup>](#),
- [Emanuele Montomoli<sup>2,3,6</sup>](#) &
- [Rino Rappuoli](#) [ORCID: orcid.org/0000-0002-8827-254X<sup>1,7</sup>](#)

[Nature](#) volume 600, pages 530–535 (2021)

- 32k Accesses
- 231 Altmetric
- [Metrics details](#)

## Subjects

- [Infectious diseases](#)
- [SARS-CoV-2](#)
- [Vaccines](#)

## Abstract

The emergence of SARS-CoV-2 variants is jeopardizing the effectiveness of current vaccines and limiting the application of monoclonal antibody-based therapy for COVID-19 (refs. [1,2](#)). Here we analysed the memory B cells of five naive and five convalescent people vaccinated with the BNT162b2 mRNA vaccine to investigate the nature of the B cell and antibody response at the single-cell level. Almost 6,000 cells were sorted, over 3,000 cells produced monoclonal antibodies against the spike protein and more than 400 cells neutralized the original SARS-CoV-2 virus first identified in Wuhan, China. The B.1.351 (Beta) and B.1.1.248 (Gamma) variants escaped almost 70% of these antibodies, while a much smaller portion was impacted by the B.1.1.7 (Alpha) and B.1.617.2 (Delta) variants. The overall loss of neutralization was always significantly higher in the antibodies from naive people. In part, this was due to the IGHV2-5;IGHJ4-1 germline, which was found only in people who were convalescent and generated potent and broadly neutralizing antibodies. Our data suggest that people who are seropositive following infection or primary vaccination will produce antibodies with increased potency and breadth and will be able to better control emerging SARS-CoV-2 variants.

[Download PDF](#)

## Main

Twenty months after the beginning of the COVID-19 pandemic, with 252 million people infected, 5 million deaths and 7.2 billion vaccine doses administered, the world is still struggling to control the virus. In most developed countries, vaccines have vastly reduced severe disease, hospitalization and deaths, but they have not been able to control the infections that are fuelled by new and more infectious variants. A large number of studies so far have shown that protection from infection is linked to the production of neutralizing antibodies against the spike (S) protein of the virus [3,4,5,6](#). This is a metastable, trimeric class 1 fusion glycoprotein, composed of the S1 and S2 subunits, and mediates virus entry, changing from a prefusion to postfusion conformation after binding to the human angiotensin-converting enzyme 2 (ACE2) receptor and heparan sulfates on the host cells [7](#). Potent neutralizing antibodies recognize the S1 subunit of each monomer, which includes the receptor-binding domain (RBD) and N-terminal domain (NTD) immunodominant sites [8](#). The large majority of neutralizing antibodies bind to the receptor-binding motif, within the RBD,

and a smaller fraction targets the NTD<sup>5,9</sup>. Neutralizing antibodies against the S2 subunit have been described; however, they have very low potency<sup>5,10</sup>. Neutralizing antibodies generated after infection derive in large part from germline IGHV3-53 and the closely related IGHV3-66 with very few somatic mutations<sup>11,12</sup>. From June 2020, the virus started to generate mutations that allowed the virus to evade neutralizing antibodies, to become more infectious, or both. Some of the mutant viruses completely replaced the original SARS-CoV-2 first detected in Wuhan, China. The most successful variant viruses are B.1.1.7 (Alpha), B.1.351 (Beta), B.1.1.248 (Gamma) and B.1.617.2 (Delta), which have been named variants of concern (VoCs)<sup>13</sup>. The Delta variant is currently spreading across the globe and causing large concerns also in fully vaccinated populations. It is therefore imperative to understand the molecular mechanisms of the immune response to vaccination to design better vaccines and vaccination policies. Several investigators have shown that vaccination of people who are convalescent can yield neutralizing antibodies that can be up to a thousand-fold higher than those induced by infection or vaccination, suggesting that one way of controlling the pandemic may be the induction of a hybrid immunity-like response using a third booster dose<sup>14,15,16,17,18</sup>. At the single-cell level, here we compared the nature of the neutralizing antibody response against the original virus first detected in Wuhan and the VoCs in naive and convalescent participants who were immunized with the BNT162b2 mRNA vaccine.

## B cell response in COVID-19 vaccinees

We enrolled ten donors who were vaccinated with the BNT162b2 mRNA vaccine: five of them were healthy people who were naive to SARS-CoV-2 infection at vaccination (seronegative) and the other five had recovered from SARS-CoV-2 infection before vaccination (seropositive). Participant details are summarized in Extended Data Table 1. Blood collection occurred at an average of 48 and 21 days after the last vaccination dose for participants who were seronegative and seropositive, respectively (Extended Data Table 1). This difference may affect the frequency of circulating B cells and the serum activity of participants who are seronegative and seropositive analysed in this study. We initially analysed the frequency of circulating B cell populations between our groups. Participants who were seropositive showed a 2.46-fold increase in S-protein-specific CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup> memory B cells compared with participants who were seronegative and an overall 10% higher count of CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup> memory B cells (Extended Data Fig. 1a–c). Conversely, participants who were seronegative showed a 2.3-fold higher frequency of CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>+</sup> memory B cells than participants who were seropositive. No differences were found in the numbers of CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>+</sup> S protein<sup>+</sup> memory B cells between the two groups assessed in this study (Extended Data Fig. 1a–c). Following the analyses of memory B cells, we characterized the polyclonal response of these donors by testing

their binding response to the S protein trimer, RBD, NTD and the S2 domain, and subsequently by testing their neutralization activity against the original SARS-CoV-2 virus first detected in Wuhan (Extended Data Fig. 2). Plasma from participants who were seropositive showed a higher binding activity to the S protein and all tested domains than plasma from participants who were seronegative (Extended Data Fig. 2a–d). In addition, participants who were seropositive showed a tenfold-higher neutralization activity against the original SARS-CoV-2 virus detected in Wuhan than in participants who were seronegative (Extended Data Fig. 2e,f).

## Isolation of neutralizing antibodies

To better characterize the B cell immune response, we single-cell-sorted antigen-specific memory B cells using the SARS-CoV-2 S protein antigen identified in Wuhan as bait, which was encoded by the mRNA vaccine. The single-cell sorting strategy was performed as previously described<sup>5</sup>. In brief, the prefusion S protein trimer-specific (S protein<sup>+</sup>), class-switched memory B cells (CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup>) were single-cell-sorted and then incubated for 2 weeks to naturally produce and release monoclonal antibodies (mAbs) into the supernatant. A total of 2,352 and 3,532 S protein<sup>+</sup> memory B cells were sorted from vaccinees who were seronegative and seropositive, respectively (Extended Data Table 2). Of these, 944 (40.1%) and 2,299 (65.1%), respectively, were released in the supernatant mAbs, recognizing the S protein prefusion trimer in enzyme-linked immunosorbent assay (ELISA) (Fig. 1a, Extended Data Table 2). These mAbs were then tested in a cytopathic effect-based microneutralization assay (CPE-MN) with the original live SARS-CoV-2 virus detected in Wuhan at a single point dilution (1:5) to identify SARS-CoV-2 neutralizing human monoclonal antibodies (nAbs). This first screening identified a total of 411 nAbs, of which 71 derived from participants who were seronegative and 340 were from participants who were seropositive (Fig. 1b, Extended Data Table 2). Overall, the fraction of S-protein-specific B cells producing nAbs were 7.5% for participants who were seronegative and 14.8% for participants who were seropositive. Following this first screening, all nAbs that were able to neutralize the SARS-CoV-2 virus detected in Wuhan were tested by CPE-MN against major VoCs, including B.1.1.7 (Alpha), B.1.351 (Beta) and B.1.1.248 (Gamma) to understand the breadth of neutralization of nAbs elicited by the BNT162b2 mRNA vaccine. At the time of this assessment the B.1.617.2 (Delta) variant had not yet spread globally and therefore was not available for screening. Participants who were seropositive had an overall higher percentage of nAbs neutralizing the VoCs than participants who were seronegative. The average frequency of nAbs from participants who were seropositive neutralizing the Alpha, Beta and Gamma variants was 80.6% ( $n = 274$ ), 39.4% ( $n = 134$ ) and 62.0% ( $n = 211$ ), respectively, compared with 70.4% ( $n = 50$ ), 22.5% ( $n = 16$ ) and 43.6% ( $n = 31$ ), respectively, in participants who were seronegative (Fig. 1c, Extended Data Table 2).

**Fig. 1: Identification of cross-neutralizing SARS-CoV-2 S protein-specific nAbs.**

 figure1

**a**, The graph shows supernatants that were tested for binding to the SARS-CoV-2 S protein antigen first detected in Wuhan, China. The threshold of positivity was set as two times the value of the blank (dotted line). The dark blue and dark red dots represent mAbs that bind to the S protein for vaccinees who were seronegative and seropositive, respectively. The light blue and light red dots represent mAbs that do not bind to the S protein for vaccinees who were seronegative and seropositive, respectively. OD, optical density. **b**, The bar graph shows the percentage of not-neutralizing antibodies (grey), nAbs from individuals who were seronegative (dark blue) and nAbs for individuals who were seropositive (dark red). The total number (*n*) of antibodies tested per individual is shown on the top of each bar in **a**, **b**. **c**, The graphs show the fold-change percentage of nAbs in individuals who were seronegative (left) and seropositive (right) against the Alpha, Beta and Gamma VoCs compared with the original SARS-CoV-2 virus detected in Wuhan. The heat maps show the overall percentage of the SARS-CoV-2 nAbs detected in Wuhan that are able to neutralize the tested VoCs.

## [Source data](#)

### Potency and breadth against variants

To better characterize and understand the potency and breadth of coverage of all nAbs against the SARS-CoV-2 virus detected in Wuhan, we aimed to express all the 411 nAbs previously identified as IgG1. We were able to recover and express 276 antibodies for further characterization, 224 (89.8%) from participants who were seropositive and 52 (10.2%) from participants who were seronegative. Initially, antibodies were tested for binding against the RBD, NTD and the S2 domain of the original SARS-CoV-2 Sprotein identified in Wuhan. Overall, no major differences were observed in nAbs that recognized the RBD and NTD, whereas nAbs that were able to bind to the S protein only in its trimeric conformation (that is, not able to bind single domains) were almost threefold higher in participants who were seronegative than in participants who were seropositive (Extended Data Fig. 3). None of the tested nAbs targeted the S2 domain. nAbs were then tested by CPE-MN in serial dilution to evaluate their 100% inhibitory concentration ( $IC_{100}$ ) against the SARS-CoV-2 virus detected in Wuhan and the VoCs. At this stage of the study, the B.1.617.2 (Delta) virus had spread globally, and we were able to obtain the live virus for our experiments. Overall, nAbs isolated from vaccinees who were seropositive had a significantly higher potency than those isolated from vaccinees who were seronegative. The  $IC_{100}$  geometric mean in participants who were seropositive was 2.87-fold, 2.17-fold, 1.17-fold, 1.43-fold and 1.92-fold lower than in participants who were seronegative for the virus detected in Wuhan, and the Alpha, Beta, Gamma and Delta VoCs, respectively (Fig. 2). In addition, a bigger fraction of nAbs from participants who were seropositive retained the ability to neutralize the VoCs. Indeed, when nAbs were individually tested against all VoCs, the ability to neutralize the Alpha, Beta, Gamma and Delta variants was lost by 14%, 61%, 61% and 29% of the antibodies from participants who were seropositive versus 32%, 78%, 75% and 46% of those from participants who were seronegative, respectively (Fig. 2). Finally, a major difference between participants who were seronegative and seropositive was found in the class of nAbs with medium/high potency ( $IC_{100}$  of 11–100 ng ml<sup>-1</sup> and 101–1,000 ng ml<sup>-1</sup>) against all variants. Indeed, nAbs in these ranges from participants who were seropositive constitute 71.0%, 62.5%, 23.7%, 22.8% and 53.1% of the whole nAb repertoire, whereas nAbs from seronegative donors were 48.1%, 38.5%, 17.3%, 17.3% and 34.6% against the SARS-CoV-2 virus detected in Wuhan and the Alpha, Beta, Gamma and Delta VoCs respectively (Fig. 2).

**Fig. 2: Potency and breadth of neutralization of nAbs against SARS-CoV-2 and VoCs.**

---

 **figure2**

**a–e**, Scatter dot charts show the neutralization potency, reported as  $IC_{100}$  ( $\text{ng ml}^{-1}$ ), of nAbs tested against the original SARS-CoV-2 virus first detected in Wuhan (**a**) and the B.1.1.7 (**b**), B.1.351 (**c**), B.1.1.248 (**d**) and B.1.617.2 (**e**) VoCs. The number and percentage of nAbs from individuals who were seronegative versus seropositive, fold change, neutralization  $IC_{100}$  geometric mean (black lines, blue and red bars) and statistical significance are denoted on each graph. A non-parametric Mann–Whitney *t*-test was used to evaluate statistical significances between groups. Two-tailed *P* value significances are shown as \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001. NS, not significant. **f**, The table shows the  $IC_{100}$  geometric mean (GM) of all nAbs pulled together from each group against all SARS-CoV-2 viruses tested. Technical duplicates were performed for each experiment.

[Source data](#)

## Functional gene repertoire

The analysis of the immunoglobulin G heavy chain variable (IGHV) and joining (IGHJ) gene rearrangements of 58 and 278 sequences recovered from participants who were seronegative and seropositive, respectively, showed that they use a broad range of germlines and share the most abundant germlines. In particular, both groups predominantly used the IGHV1-69;IGHJ4-1 and IGHV3-53;IGHJ6-1 germlines, which were shared by three out five participants per group (Fig. 3a). In addition, the IGHV3-30;IGHJ6-1 and IGHV3-33;IGHJ4-1 germlines, which were more abundant in donors who were seronegative, and the IGHV1-2;IGHJ6-1 germline, which was mainly expanded in vaccinees who were seropositive, were also used with high frequency in both groups. Only the IGHV2-5;IGHJ4-1 germline was seen to be predominantly expanded only in donors who were seropositive (Fig. 3a). Despite the

fact that selected germlines were boosted following vaccination, no major clonal families were identified, and the biggest family observed contained only four antibodies. To better characterize these predominant gene families, we evaluated their neutralization potency and breadth against SARS-CoV-2 and VoCs. In this analysis, we could not evaluate IGHV3-33;IGHJ4-1 nAbs, as only three of these antibodies were expressed, but we included the IGHV3-53 closely related family IGHV3-66;IGHJ4-1, as this family was previously described to be mainly involved in the neutralization of the SARS-CoV-2 virus<sup>11,19</sup>. A large part of nAbs deriving from these predominant germlines had a very broad range of neutralization potency against the original SARS-CoV-2 virus detected in Wuhan, with the IC<sub>100</sub> spanning from less than 10 to over 10,000 ng ml<sup>-1</sup> (Fig. 3b–g). However, many of them lost the ability to neutralize SARS-CoV-2 VoCs. The loss of neutralizing activity occurred for most germlines and it was moderate against the Alpha and Delta variants, whereas the loss was marked against the Beta and Gamma variants (Fig. 3b–g). A notable exception was the IGHV2-5;IGHJ4-1 germline, which was present only in nAbs of participants who were seropositive, that showed potent antibodies able to equally neutralize all SARS-CoV-2 VoCs (Fig. 3d). Finally, we evaluated the CDRH3 length and V-gene somatic hypermutation levels for all nAbs retrieved from participants who were seronegative and seropositive and for predominant germlines. Overall, the two groups show a similar average CDRH3 length (15.0 amino acids and 15.1 amino acids for participants who were seronegative and seropositive, respectively); however, participants who were seropositive showed almost twofold-higher V-gene mutation levels than participants who were seronegative (Extended Data Fig. 4). As for predominant gene-derived nAbs, we observed heterogenous CDRH3 length, with the only exception of IGHV3-53;IGHJ6-1 nAbs, and higher V-gene mutation levels in predominant germlines from participants who were seropositive than in germlines from participants who were seronegative (Extended Data Fig. 5).

**Fig. 3: Repertoire analyses and functional characterization of predominant gene-derived nAbs.**

---

 **figure3**

**a**, The graph shows the IGHV-J rearrangement frequencies between vaccinees who were seronegative and seropositive (top), and the frequency within seronegative (middle) and seropositive (bottom) participants. **b–g**, The graphs show the neutralization potency ( $IC_{100}$ ) of predominant gene-derived nAbs from the IGHV1-2;IGHJ6-1 (**b**), IGHV1-69;IGHJ4-1 (**c**), IGHV2-5;IGHJ4-1 (**d**), IGHV3-30;IGHJ6-1 (**e**), IGHV3-53;IGHJ6-1 (**f**) and IGHV3-66;IGHJ4-1 (**g**) families, against the original SARS-CoV-2 virus first detected in Wuhan and the B.1.1.7, B.1.351, B.1.1.248 and B.1.617.2 VoCs.

[Source data](#)

## S protein epitope mapping

To map the regions of the S protein recognized by the identified nAbs, we used a competition assay with four known antibodies: J08, which targets the top loop of the receptor-binding motif<sup>5</sup>; S309, which binds to the RBD but outside of the RMB region<sup>20</sup>; 4A8, which recognizes the NTD<sup>21</sup>; and L19, which binds to the S2 domain<sup>5</sup> (Extended Data Fig. 6). The nAbs identified in this study were pre-incubated with the original SARS-CoV-2 S protein detected in Wuhan, and subsequently the four nAbs labelled with different fluorophores were added as a single mix. For one of the four fluorescently labelled nAbs, 50% signal reduction was used as a threshold for positive competition. The vast majority of nAbs from both seronegative (50.0%;  $n = 26$ ) and seropositive (51.3%;  $n = 115$ ) vaccinees competed with J08 (Extended Data Fig. 7a, Extended Data Table 3). For vaccinees who were seronegative, the second most abundant population was composed of nAbs that did not compete with any of the four fluorescently labelled nAbs (25.0%;  $n = 13$ ), followed by nAbs targeting the NTD (17.3%;  $n = 9$ ). As for vaccinees who were seropositive, the second most abundant population was composed of nAbs that competed with S309 (21.4%;  $n = 48$ ), followed by nAbs that competed with 4A8 (15.6%;  $n = 35$ ) and not-competing nAbs (11.6%;  $n = 26$ ). None of our nAbs competed with the S2-targeting antibody L19 (Extended Data Fig. 7a, Extended Data Table 3). nAbs that competed with J08, which are likely to bind to the receptor-binding motif, derived from several germlines, including the predominant IGHV3-53;IGHJ6-1 (10.6%;  $n = 14$ ), IGHV1-69;IGHJ4-1 (8.3%;  $n = 11$ ) and IGHV1-2;IGHJ6-1 (6.8%;  $n = 9$ ) germlines (Extended Data Fig. 7b). By contrast, those that competed with S309 derived mostly from the IGHV2-5;IGHJ4-1 germline (13.7%;  $n = 7$ ), which were isolated exclusively from vaccinees who were seropositive (Extended Data Fig. 7c). As for NTD-directed nAbs, the non-predominant gene family IGHV1-24;IGHJ6-1 was the most abundant, confirming what was reported in previous studies<sup>22</sup> (Extended Data Fig. 7d). Finally, for nAbs that did not compete with any of the known antibodies used in our competition assay, the non-predominant gene families IGHV1-69;IGHJ3-1 (9.7%;  $n = 3$ ) and IGHV1-69;IGHJ6-1 (9.7%;  $n = 3$ ) were the most abundant (Extended Data Fig. 7e).

## Discussion

Our study analysed the repertoire of B cells producing neutralizing antibodies following vaccination of naive and previously infected people at the single-cell level. The most important conclusion from this work is that people who are previously exposed to SARS-CoV-2 infection respond to vaccination with more B-cell-producing antibodies that are not susceptible to escape variants and that have higher neutralization potency. This can be explained in part by the increased number of somatic mutations and by the fact that participants who are seropositive expand potent antibodies derived from the IGHV2-5;IGHJ4-1 germline, which were not described in naive vaccinees<sup>18</sup>. One limitation of our study is that we did not include people who

received a third booster dose of vaccine. Despite this limitation, we believe that our conclusions are likely to be extendable to people who are seronegative, as a third vaccine dose could lead to a hybrid immunity-like response as neutralizing antibodies following infection and vaccination derive mostly from the same immunodominant germlines<sup>11,12,17,18,19</sup>. Our analysis suggests that a booster dose of vaccine will increase the frequency of memory B cells producing potent neutralizing antibodies that are not susceptible to escape variants and will allow better control of this pandemic. The massive variant escape from predominant germlines, such as IGHV3-53, IGHV3-66, IGHV3-30 and IGHV1-69, and the presence of antibodies deriving from the IGHV2-5 germline that are resistant to variants, suggest that the design of vaccines that preferentially promote or avoid the expansion of selected germlines can generate broad protection against SARS-CoV-2 variants. Germline-targeting vaccination, which has been pioneered in the HIV field<sup>23,24</sup>, may be a promising strategy to fight the COVID-19 pandemic.

## Methods

### Enrolment of COVID-19 vaccinees and human sample collection

This work results from a collaboration with the Azienda Ospedaliera Universitaria Senese, Siena (IT) that provided samples from donors vaccinated against COVID-19, of both sexes, who gave their written consent. The study was approved by the Comitato Etico di Area Vasta Sud Est (CEAVSE) ethics committees (Parere 17065 in Siena) and conducted according to good clinical practice in accordance with the declaration of Helsinki (European Council 2001, US Code of Federal Regulations, ICH 1997). This study was unblinded and not randomized. No statistical methods were used to predetermine sample size.

### Single-cell sorting of SARS-CoV-2 S protein<sup>+</sup> memory B cells from COVID-19 vaccinees

Peripheral blood mononuclear cells (PBMCs) and the single-cell sorting strategy were performed as previously described<sup>5</sup>. In brief, PBMCs were isolated from heparin-treated whole blood by density gradient centrifugation (Ficoll-Paque PREMIUM, Sigma-Aldrich). After separation, PBMCs were stained with Live/Dead Fixable Aqua (Invitrogen; Thermo Scientific) diluted 1:500 at room temperature. After 20 min of incubation, cells were washed with PBS and unspecific bindings were saturated with 20% normal rabbit serum (Life Technologies). Following 20 min of incubation at 4 °C, cells were washed with PBS and stained with SARS-CoV-2 S protein labelled with Strep-Tactin XT DY-488 (2-1562-050, iba-lifesciences) for 30 min at 4 °C. After incubation, the following staining mix was used CD19 V421 (1:320; 562440, BD),

IgM PerCP-Cy5.5 (1:50; 561285, BD), CD27 PE (1:30; 340425, BD), IgD-A700 (1:15; 561302, BD), CD3 PE-Cy7 (1:100; 300420, BioLegend), CD14 PE-Cy7 (1:320; 301814, BioLegend), CD56 PE-Cy7 (1:80; 318318, BioLegend) and cells were incubated at 4 °C for an additional 30 min. Stained memory B cells were single-cell-sorted with a BD FACS Aria III (BD Biosciences) into 384-well plates containing 3T3-CD40L feeder cells and were incubated with IL-2 and IL-21 for 14 days as previously described<sup>25</sup>.

### **ELISA assay with SARS-CoV-2 S protein prefusion trimer**

mAbs and plasma binding specificity against the S protein trimer was detected by ELISA as previously described<sup>5</sup>. In brief, 384-well plates (microplate clear, Greiner Bio-one) were coated with 3 µg/ml of streptavidin (Thermo Fisher) diluted in carbonate-bicarbonate buffer (E107, Bethyl Laboratories) and incubated at room temperature overnight. The next day, plates were incubated for 1 h at room temperature with 3 µg/ml of SARS-CoV-2 S protein diluted in PBS. Plates were then saturated with 50 µl per well of blocking buffer (phosphate-buffered saline and 1% BSA) for 1 h at 37 °C. After blocking, 25 µl per well of mAbs diluted 1:5 in sample buffer (phosphate-buffered saline, 1% BSA and 0.05% Tween-20) was added to the plates and was incubated at 37 °C. Plasma samples derived from vaccinees were tested (starting dilution of 1:10; step dilution of 1:2 in sample buffer) in a final volume of 25 µl per well and were incubated at 37 °C. After 1 h of incubation, 25 µl per well of alkaline phosphatase-conjugated goat anti-human IgG and IgA (Southern Biotech) diluted 1:2,000 in sample buffer was added. Finally, S protein binding was detected using 25 µl per well of PNPP (p-nitrophenyl phosphate; Thermo Fisher) and the reaction was measured at a wavelength of 405 nm by the Varioskan Lux Reader (Thermo Fisher Scientific). After each incubation step, plates were washed three times with 100 µl per well of washing buffer (phosphate-buffered saline and 0.05% Tween-20). Sample buffer was used as a blank and the threshold for sample positivity was set at twofold the optical density (OD) of the blank. Technical duplicates were performed for mAbs and technical triplicates were performed for sera samples.

### **ELISA assay with RBD, NTD and S2 subunits**

Identification of mAbs and plasma screening of vaccinees against RBD, NTD or S2 SARS-CoV-2 protein were performed by ELISA. In brief, 3 µg/ml of RBD, NTD or S2 SARS-CoV-2 protein diluted in carbonate-bicarbonate buffer (E107, Bethyl Laboratories) was coated in 384-well plates (microplate clear, Greiner Bio-one). After overnight incubation at 4 °C, plates were washed three times with washing buffer (phosphate-buffered saline and 0.05% Tween-20) and blocked with 50 µl per well of blocking buffer (phosphate-buffered saline and 1% BSA) for 1 h at 37 °C. After washing, plates were incubated 1 h at 37 °C with mAbs diluted 1:5 in sample buffer

(phosphate-buffered saline, 1% BSA and 0.05% Tween-20) or with plasma at a starting dilution of 1:10 and step diluted of 1:2 in sample buffer. Wells with no sample added were considered blank controls. Anti-human IgG-peroxidase antibody (Fab specific) produced in goat (Sigma) diluted 1:45,000 in sample buffer was then added and samples were incubated for 1 h at 37 °C. Plates were then washed, incubated with TMB substrate (Sigma) for 15 min before adding the stop solution ( $H_2SO_4$  0.2 M). The OD values were identified using the Varioskan Lux Reader (Thermo Fisher Scientific) at 450 nm. Each condition was tested in triplicate and samples tested were considered positive if the OD value was twofold the blank.

### **Flow cytometry-based competition assay**

To classify mAb candidates on the basis of their interaction with S epitopes, we performed a flow cytometry-based competition assay. In detail, magnetic beads (Dynabeads His-Tag, Invitrogen) were coated with histidine-tagged S protein according to the manufacturer's instructions. Then, 20 µg/ml of coated S protein beads were pre-incubated with unlabelled nAb candidates diluted 1:2 in PBS for 40 min at room temperature. After incubation, the mix beads–antibodies was washed with 100 µl of 1% PBS-BSA. Then, to analyse epitope competition, mAbs that are able to bind to the RBD (J08 and S309), NTD (4A8) or S2 domain (L19) of the S protein were labelled with four different fluorophores (Alexa Fluor 647, 488, 594 and 405) using the Alexa Fluor NHS Ester kit (Thermo Scientific), were mixed and incubated with S protein beads. Following 40 min of incubation at room temperature, the mix beads–antibodies was washed with PBS, resuspended in 150 µl of 1% PBS-BSA and analysed using the BD LSR II flow cytometer (Becton Dickinson). Beads with or without S protein incubated with labelled antibodies mix were used as positive and negative controls, respectively. FACSDiva Software (version 9) was used for data acquisition and analysis was performed using FlowJo (version 10).

### **SARS-CoV-2 authentic viruses neutralization assay**

All SARS-CoV-2 authentic virus neutralization assays were performed in the biosafety level 3 (BSL3) laboratories at Toscana Life Sciences in Siena (Italy) and Vismederi Srl, Siena (Italy). BSL3 laboratories are approved by a certified biosafety professional and are inspected every year by local authorities. To evaluate the neutralization activity of identified nAbs against SARS-CoV-2 and all VoCs and to evaluate the breadth of neutralization of this antibody, CPE-MN was performed<sup>5</sup>. In brief, for the CPE-based neutralization assay, we co-incubated J08 with a SARS-CoV-2 viral solution containing 100 median tissue culture infectious dose (100 TCID<sub>50</sub>) of virus, and after 1 h of incubation at 37 °C, 5% CO<sub>2</sub>. The mixture was then added to the wells of a 96-well plate containing a sub-confluent Vero E6 cell monolayer. Plates were

incubated for 3–4 days at 37 °C in a humidified environment with 5% CO<sub>2</sub>, then examined for CPE by means of an inverted optical microscope by two independent operators. All nAbs were tested at a starting dilution of 1:5 and the IC<sub>100</sub> was evaluated based on their initial concentration, while plasma samples were tested starting at a 1:10 dilution. Both nAbs and plasma samples were then step diluted 1:2. Technical duplicates were performed for both nAbs and plasma samples. In each plate, positive and negative controls were used as previously described<sup>5</sup>.

### SARS-CoV-2 virus variants CPE-MN neutralization assay

The SARS-CoV-2 viruses used to perform the CPE-MN neutralization assay were the original SARS-CoV-2 virus first detected in Wuhan (SARS-CoV-2/INMI1-Isolate/2020/Italy: MT066156), SARS-CoV-2 B.1.1.7 (INMI GISAID accession number: EPI\_ISL\_736997), SARS-CoV-2 B.1.351 (EVAg Cod: 014V-04058), B.1.1.248 (EVAg CoD: 014V-04089) and B.1.617.2 (GISAID ID: EPI\_ISL\_2029113)<sup>26</sup>.

### Single-cell RT–PCR and Ig gene amplification and transcriptionally active PCR expression

The whole process for nAbs heavy and light chain recovery, amplification and transcriptionally active PCR (TAP) expression was performed as previously described<sup>5</sup>. In brief, 5 µl of cell lysate was mixed with 1 µl of random hexamer primers (50 ng/µl), 1 µl of dNTP-Mix (10 mM), 2 µl of 0.1 M DTT, 40 U/µl of RNase OUT, MgCl<sub>2</sub> (25 mM), 5× FS buffer and Superscript IV reverse transcriptase (Invitrogen) to perform RT–PCR. Reverse transcription (RT) reaction was performed at 42 °C for 10 min, 25 °C for 10 min, 50 °C for 60 min and 94 °C for 5 min. Two rounds of PCR were performed to obtain the heavy (VH) and light (VL) chain amplicons. All PCRs were performed in a nuclease-free water (DEPC) in a total volume of 25 µl per well. For PCR I, 4 µl of cDNA were mixed with 10 µM of VH and 10 µM of VL primer-mix, 10 mM of dNTP mix, 0.125 µl of Kapa Long Range Polymerase (Sigma), 1.5 µl of MgCl<sub>2</sub> and 5 µl of 5× Kapa Long Range Buffer. The PCR I reaction was performed at 95 °C for 3 min, 5 cycles at 95 °C for 30 s, 57 °C for 30 s, 72 °C for 30 s and 30 cycles at 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s and a final extension of 72 °C for 2 min. Nested PCR II was performed as above starting from 3.5 µl of unpurified PCR I product. PCR II products were purified by the Millipore MultiScreen PCRµ96 plate according to manufacturer's instructions and eluted in 30 µl of nuclease-free water (DEPC). As for TAP expression, vectors were initially digested using restriction enzymes AgeI, SalI and Xho as previously described and PCR II products ligated by using the Gibson Assembly NEB into 25 ng of respective human Igγ1, Igκ and Igλ expression vectors<sup>27,28</sup>. TAP reaction was

performed using 5 µl of Q5 polymerase (NEB), 5 µl of GC Enhancer (NEB), 5 µl of 5X buffer, 10 mM of dNTPs, 0.125 µl of forward/reverse primers and 3 µl of ligation product, using the following cycles: 98 °C for 2 min, 35 cycles 98 °C for 10 s, 61 °C for 20 s, 72 °C for 1 min and 72 °C for 5 min. TAP products were purified under the same PCR II conditions, quantified by the Qubit Fluorometric Quantitation assay (Invitrogen) and used for transient transfection in the Expi293F cell line following the manufacturer's instructions.

## Functional repertoire analyses

The VH and VL sequence reads of nAbs were manually curated and retrieved using CLC sequence viewer (Qiagen). Aberrant sequences were removed from the dataset. Analysed reads were saved in FASTA format and the repertoire analyses were performed using Cloanalyst

(<http://www.bu.edu/computationalimmunology/research/software/>)<sup>29,30</sup>.

## Statistical analysis

Statistical analysis was assessed with GraphPad Prism Version 8.0.2 (GraphPad Software). Non-parametric Mann–Whitney *t*-test was used to evaluate statistical significance between the two groups analysed in this study. Statistical significance was shown as \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , \*\*\*\* $P \leq 0.0001$ .

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Source data are provided with this paper. All data supporting the findings in this study are available within the article or can be obtained from the corresponding author on request. [Source data](#) are provided with this paper.

## References

1. 1.

Krause, P. R. et al. SARS-CoV-2 variants and vaccines. *N. Engl. J. Med.* **385**, 179–186 (2021).

2. 2.

Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).

3. 3.

Wajnberg, A. et al. Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science* **370**, 1227–1230 (2020).

4. 4.

Xiaojie, S., Yu, L., Lei, Y., Guang, Y. & Min, Q. Neutralizing antibodies targeting SARS-CoV-2 spike protein. *Stem Cell Res.* **50**, 102125 (2021).

5. 5.

Andreano, E. et al. Extremely potent human monoclonal antibodies from COVID-19 convalescent patients. *Cell* **184**, 1821–1835.e16 (2021).

6. 6.

Yang, Y. & Du, L. SARS-CoV-2 spike protein: a key target for eliciting persistent neutralizing antibodies. *Signal Transduct. Target. Ther.* **6**, 95 (2021).

7. 7.

Clausen, T. M., et al. SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* **183**, 1043-1057.e15 (2020).

8. 8.

Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).

9. 9.

Piccoli, L. et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).

10. 10.

Tortorici, M. A. et al. Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* **597**, 103–108 (2021).

11. 11.

Andreano, E. & Rappuoli, R. Immunodominant antibody germlines in COVID-19. *J. Exp. Med.* **218**, e20210281 (2021).

12. 12.

Yuan, M. et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119–1123 (2020).

13. 13.

Callaway, E. Coronavirus variants get Greek names — but will scientists use them? *Nature* **594**, 162 (2021).

14. 14.

Stamatatos, L. et al. mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021).

15. 15.

Goel, R. R. et al. Distinct antibody and memory B cell responses in SARS-CoV-2 naïve and recovered individuals after mRNA vaccination. *Sci. Immunol.* **6**, eabi6950 (2021).

16. 16.

Urbanowicz, R. A. et al. Two doses of the SARS-CoV-2 BNT162b2 vaccine enhances antibody responses to variants in individuals with prior SARS-CoV-2 infection. *Sci. Transl. Med.* **13**, eabj0847 (2021).

17. 17.

Crotty, S. Hybrid immunity. *Science* **372**, 1392–1393 (2021).

18. 18.

Wang, Z. et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021).

19. 19.

Zhang, Q. et al. Potent and protective IGHV3-53/3-66 public antibodies and their shared escape mutant on the spike of SARS-CoV-2. *Nat. Commun.* **12**, 4210 (2021).

20. 20.

Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).

21. 21.

Chi, X. et al. A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020).

22. 22.

Voss, W. N. et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science* **372**, 1108–1112 (2021).

23. 23.

Steichen, J. M. et al. A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science* **366**, eaax4380 (2019).

24. 24.

Havenar-Daughton, C., Abbott, R. K., Schief, W. R. & Crotty, S. When designing vaccines, consider the starting material: the human B cell repertoire. *Curr. Opin. Immunol.* **53**, 209–216 (2018).

25. 25.

Huang, J. et al. Isolation of human monoclonal antibodies from peripheral blood B cells. *Nat. Protoc.* **8**, 1907–1915 (2013).

26. 26.

Planas, D. et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat. Med.* **27**, 917–924 (2021).

27. 27.

Tiller, T. et al. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* **329**, 112–124 (2008).

28. 28.

Wardemann, H. & Busse, C. E. Expression cloning of antibodies from single human B cells. *Methods Mol. Biol.* **1956**, 105–125 (2019).

29. 29.

Kepler, T. B. Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000Res.* **2**, 103 (2013).

30. 30.

Kepler, T. B. et al. Reconstructing a B-cell clonal lineage. II. Mutation, selection, and affinity maturation. *Front. Immunol.* **5**, 170 (2014).

## Acknowledgements

This work was funded by the European Research Council (ERC) advanced grant agreement number 787552 (vAMRes). This publication was supported by funds from the ‘Centro Regionale Medicina di Precisione’ and by all of the people who answered the call to fight with us in the battle against SARS-CoV-2 with their kind donations on the platform ForFunding (<https://www.forfunding.intesasanpaolo.com/DonationPlatform-ISP/nav/progetto/id/3380>). This work was funded by COOP ITALIA Soc. Coop. This publication was supported by the European Virus Archive goes Global (EVAg) project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 653316. This publication was supported by the COVID-2020-12371817 project, which has received funding from the Italian Ministry of Health. We also acknowledge J. McLellan, for kindly providing the S protein trimer, RBD, NTD and S2 constructs; O. Schwartz, for providing the B.1.617.2 (Delta) SARS-CoV-2 variant; the nursing staff of the operative unit of the department of Medical Sciences, Infectious and Tropical Diseases Unit, Siena University Hospital, Siena, Italy; and all of the donors who are vaccinated against COVID-19 for participating to this study.

## Author information

### Affiliations

1. Monoclonal Antibody Discovery (MAD) Lab, Fondazione Toscana Life Sciences, Siena, Italy

Emanuele Andreano, Ida Paciello, Noemi Manganaro, Piero Pileri, Elisa Pantano, Valentina Abbiento, Concetta De Santi, Claudia Sala & Rino Rappuoli

2. VisMederi S.r.l, Siena, Italy

Giulia Piccini, Inesa Hyseni, Margherita Leonardi & Emanuele Montomoli

3. VisMederi Research S.r.l., Siena, Italy

Inesa Hyseni, Margherita Leonardi, Linda Benincasa, Ginevra Giglioli & Emanuele Montomoli

4. Department of Medical Sciences, Infectious and Tropical Diseases Unit, Siena University Hospital, Siena, Italy

Massimiliano Fabbiani, Ilaria Rancan, Mario Tumbarello & Francesca Montagnani

5. Department of Medical Biotechnologies, University of Siena, Siena, Italy

Ilaria Rancan, Mario Tumbarello & Francesca Montagnani

6. Department of Molecular and Developmental Medicine, University of Siena, Siena, Italy

Emanuele Montomoli

7. Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, Italy

Rino Rappuoli

## Contributions

E.A. and R.R. conceived the study. F.M., M.F., I.R. and M.T. enrolled the COVID-19 vaccinees. E.A. and I.P. performed PBMC isolation and single-cell sorting. I.P. performed ELISAs and competition assays. I.P. and N.M. recovered nAbs expressing VH and VL and antibodies. P.P. and E.A. recovered the VH and VL sequences and performed the repertoire analyses. E.P. and V.A. produced and purified the SARS-CoV-2 S protein constructs. E.A., G.P., I.H., M.L., L.B. and G.G. performed the neutralization assays in the BSL3 facilities. C.D.S. supported day-to-day laboratory

activities and management. E.A. and R.R. wrote the manuscript. E.A., I.P., G.P., N.M., P.P., I.H., M.L., E.P., V.A., L.B., G.G., C.D.S., M.F., I.R., M.T., F.M., C.S., E.M. and R.R. undertook the final revision of the manuscript. E.A., C.S., E.M. and R.R. coordinated the project.

## Corresponding author

Correspondence to [Rino Rappuoli](#).

## Ethics declarations

## Competing interests

R.R. is an employee of the GSK group of companies. E.A., I.P., N.M., P.P., E.P., C.D.S., C.S. and R.R. are listed as inventors of full-length human mAbs described in Italian patent applications no. 102020000015754 filed on 30 June 2020, 102020000018955 filed on 3 August 2020 and 102020000029969 filed on 4 December 2020, and the international patent system number PCT/IB2021/055755 filed on 28 June 2021. All patents were submitted by Fondazione Toscana Life Sciences, Siena, Italy. The remaining authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Single cell sorting and memory B cell frequencies.](#)

**a, b**, The gating strategy shows from left to right: CD19<sup>+</sup> B cells; CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>; CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup>/IgM<sup>+</sup>; CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup>Sprotein<sup>+</sup>; CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup>Sprotein<sup>+</sup> for a healthy donor (used as negative control for S protein staining) and a vaccinated subject. **c**, The graph shows the frequency of CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup> and IgM<sup>+</sup> in seronegative (n=5) and seropositive donors (n=5). **d**, The graph shows the frequency of CD19<sup>+</sup>CD27<sup>+</sup>IgD<sup>-</sup>IgM<sup>-</sup> and IgM<sup>+</sup> able to bind the SARS-CoV-2 S protein trimer (S protein<sup>+</sup>) in seronegative (n=5) and seropositive (n=5) donors.

Geometric mean and standard deviation are denoted on the graphs. A nonparametric Mann–Whitney t test was used to evaluate statistical significances between groups. Two-tailed p-value significances are shown as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ . e, The table summarizes the frequencies of the cell population above described for all subjects enrolled in our study.

[Source data](#)

### **Extended Data Fig. 2 Plasma response of COVID-19 vaccinees.**

**a–d**, Graphs show the ability of plasma samples from seronegative and seropositive vaccinees to bind the S protein trimer, RBD, NTD and S2 domain. Mean and standard deviation are denoted on each graph. Technical triplicates were performed for each experiment. e, The graph shows the neutralizing activity of plasma samples against the original Wuhan SARS-CoV-2 virus. Technical duplicates were performed for each experiment. f, The table summarizes the 100% inhibitory dilution ( $ID_{100}$ ) of each COVID-19 vaccinee and the geometric mean for seronegative and seropositive donors.

[Source data](#)

### **Extended Data Fig. 3 RBD and NTD binding distribution of nAbs.**

The graph shows the percentage of antibodies that bind specifically the RBD (light orange) or the NTD (cyan) or that did not bind single domains but recognized exclusively the S protein in its trimetric conformation (gray). The number (n) of tested nAbs per donor is reported on top of each bar. Technical duplicates were performed for each experiment.

[Source data](#)

### **Extended Data Fig. 4 Heavy chain CDR3 length and somatic hypermutation levels in seronegative and seropositive vaccinees.**

**a**, The graph shows the heavy chain CDR3 length represented in amino acids (aa). **b**, The graph shows the overall somatic hypermutation level of nAbs isolated from seronegative and seropositive vaccinees. Geometric mean and standard deviation are denoted on the graphs. A nonparametric Mann–Whitney t test was used to evaluate statistical significances between groups. Two-tailed p-value significances are shown as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ .

[Source data](#)

## **Extended Data Fig. 5 Heavy chain CDR3 length and somatic hypermutation levels of predominant gene derived nAbs.**

**a–f**, Graphs show the amino acidic heavy chain CDR3 length (left panel) and the somatic hypermutation level (right panel) of nAbs derived from the IGHV1-2;IGHJ6-1 ( $n = 13$ ), IGHV1-69;IGHJ4-1 ( $n = 33$ ), IGHV2-5;IGHJ4-1 ( $n = 7$ ), IGHV3-30;IGHJ6-1 ( $n = 10$ ), IGHV3-53;IGHJ6-1 ( $n = 15$ ) and IGHV3-66;IGHJ4-1 ( $n = 9$ ) gene families. Geometric mean and standard deviation are denoted on the graphs. A nonparametric Mann–Whitney t test was used to evaluate statistical significances between groups. Two-tailed p-value significances are shown as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ .

[Source data](#)

## **Extended Data Fig. 6 Epitope binning assay.**

**a**, Schematic representation of the epitopes recognized by J08 (dark red), S309 (orange), 4A8 (dark blue) and L19 (gold), mAbs on the S protein surface. **b–e**, Representative cytometer peaks per each of the four mAbs used for the competition assay. Positive (beads conjugated with only primary labeled antibody) and negative (un-conjugated beads) controls are shown as green and gray peaks, respectively.

[Source data](#)

## **Extended Data Fig. 7 Epitope binning and genetic characterization of competing nAbs.**

**a**, The bar graph shows the percentage (%) of nAbs competing with J08 (dark red), S309 (orange), 4A8 (dark blue) and L19 (gold), or antibodies that did not compete with any of the previous mAbs (gray). A schematic representation of J08, S309, 4A8 and L19 epitopes on the S protein surface is shown on the left side of the panel. **b–e**, Graphs show the IGHV-J rearrangement percentage for nAbs that competed against J08, S309, 4A8, or that did not compete with any of these mAbs. The total number (n) of competing nAbs per group is shown on top of each graph.

[Source data](#)

**Extended Data Table 1 Clinical details of COVID-19 vaccinees**

**Extended Data Table 2 Summary of B cell frequencies and antibodies of COVID-19 vaccinees**

**Extended Data Table 3 Competition assay summary**

# **Supplementary information**

## **Reporting Summary**

## **Peer Review File**

## **Source data**

**Source Data Fig. 1**

**Source Data Fig. 2**

**Source Data Fig. 3**

**Source Data Extended Data Fig. 1**

**Source Data Extended Data Fig. 2**

**Source Data Extended Data Fig. 3**

**Source Data Extended Data Fig. 4**

**Source Data Extended Data Fig. 5**

**Source Data Extended Data Fig. 6**

**Source Data Extended Data Fig. 7**

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the

permitted use, you will need to obtain permission directly from the copyright holder.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## [Reprints and Permissions](#)

## About this article

### Cite this article

Andreano, E., Paciello, I., Piccini, G. *et al.* Hybrid immunity improves B cells and antibodies against SARS-CoV-2 variants. *Nature* **600**, 530–535 (2021).  
<https://doi.org/10.1038/s41586-021-04117-7>

- Received: 19 August 2021
- Accepted: 08 October 2021
- Published: 20 October 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04117-7>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04117-7>

- Article
- [Published: 24 November 2021](#)

# A multi-scale map of cell structure fusing protein images and interactions

- [Yue Qin](#) ORCID: [orcid.org/0000-0002-0496-7134<sup>1,2</sup>](https://orcid.org/0000-0002-0496-7134),
- [Edward L. Huttlin](#) ORCID: [orcid.org/0000-0002-1822-1173<sup>3</sup>](https://orcid.org/0000-0002-1822-1173),
- [Casper F. Winsnes](#) ORCID: [orcid.org/0000-0002-0028-5865<sup>4</sup>](https://orcid.org/0000-0002-0028-5865),
- [Maya L. Gosztyla](#) ORCID: [orcid.org/0000-0003-4758-4800<sup>5,6,7</sup>](https://orcid.org/0000-0003-4758-4800),
- [Ludivine Wacheul](#) ORCID: [orcid.org/0000-0002-0443-8865<sup>8</sup>](https://orcid.org/0000-0002-0443-8865),
- [Marcus R. Kelly](#)<sup>1</sup>,
- [Steven M. Blue](#) ORCID: [orcid.org/0000-0003-0279-6690<sup>5,6,7</sup>](https://orcid.org/0000-0003-0279-6690),
- [Fan Zheng](#)<sup>1</sup>,
- [Michael Chen](#)<sup>1</sup>,
- [Leah V. Schaffer](#)<sup>1</sup>,
- [Katherine Licon](#)<sup>1</sup>,
- [Anna Bäckström](#)<sup>4</sup>,
- [Laura Pontano Vaites](#)<sup>3</sup>,
- [John J. Lee](#) ORCID: [orcid.org/0000-0002-7967-3384<sup>1</sup>](https://orcid.org/0000-0002-7967-3384),
- [Wei Ouyang](#) ORCID: [orcid.org/0000-0002-0291-926X<sup>4</sup>](https://orcid.org/0000-0002-0291-926X),
- [Sophie N. Liu](#) ORCID: [orcid.org/0000-0001-7934-8605<sup>1</sup>](https://orcid.org/0000-0001-7934-8605),
- [Tian Zhang](#)<sup>3</sup>,
- [Erica Silva](#)<sup>1</sup>,
- [Jisoo Park](#)<sup>1</sup>,
- [Adriana Pitea](#)<sup>1</sup>,
- [Jason F. Kreisberg](#) ORCID: [orcid.org/0000-0002-3616-079X<sup>1</sup>](https://orcid.org/0000-0002-3616-079X),
- [Steven P. Gygi](#) ORCID: [orcid.org/0000-0001-7626-0034<sup>3</sup>](https://orcid.org/0000-0001-7626-0034),
- [Jianzhu Ma](#) ORCID: [orcid.org/0000-0002-8236-6609<sup>9</sup>](https://orcid.org/0000-0002-8236-6609),

- [J. Wade Harper](#) ORCID: [orcid.org/0000-0002-6944-7236<sup>3</sup>](https://orcid.org/0000-0002-6944-7236),
- [Gene W. Yeo](#) ORCID: [orcid.org/0000-0002-0799-6037<sup>2,5,6,7</sup>](https://orcid.org/0000-0002-0799-6037),
- [Denis L. J. Lafontaine](#) ORCID: [orcid.org/0000-0001-7295-6288<sup>8</sup>](https://orcid.org/0000-0001-7295-6288),
- [Emma Lundberg](#) ORCID: [orcid.org/0000-0001-7034-0850<sup>4,10,11</sup>](https://orcid.org/0000-0001-7034-0850) &
- [Trey Ideker](#) ORCID: [orcid.org/0000-0002-1708-8454<sup>1,2,7,12,13</sup>](https://orcid.org/0000-0002-1708-8454)

[\*Nature\*](#) volume **600**, pages 536–542 (2021)

- 12k Accesses
- 366 Altmetric
- [Metrics details](#)

## Subjects

- [Computational models](#)
- [Data integration](#)
- [Machine learning](#)
- [Network topology](#)
- [Proteome informatics](#)

## Abstract

The cell is a multi-scale structure with modular organization across at least four orders of magnitude<sup>1</sup>. Two central approaches for mapping this structure—protein fluorescent imaging and protein biophysical association—each generate extensive datasets, but of distinct qualities and resolutions that are typically treated separately<sup>2,3</sup>. Here we integrate immunofluorescence images in the Human Protein Atlas<sup>4</sup> with affinity purifications in BioPlex<sup>5</sup> to create a unified hierarchical map of human cell architecture. Integration is achieved by configuring each approach as a general measure of protein distance, then calibrating the two measures using machine learning. The map, known as the multi-scale integrated cell (MuSIC 1.0), resolves 69 subcellular systems, of which approximately half are to our knowledge undocumented. Accordingly, we perform 134

additional affinity purifications and validate subunit associations for the majority of systems. The map reveals a pre-ribosomal RNA processing assembly and accessory factors, which we show govern rRNA maturation, and functional roles for SRRM1 and FAM120C in chromatin and RPS3A in splicing. By integration across scales, MuSIC increases the resolution of imaging while giving protein interactions a spatial dimension, paving the way to incorporate diverse types of data in proteome-wide cell maps.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

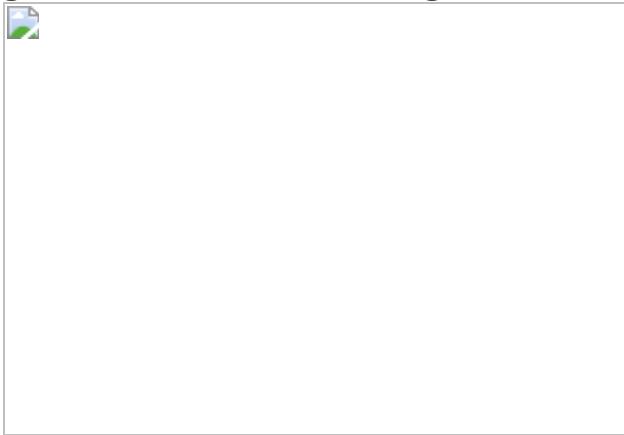
## Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

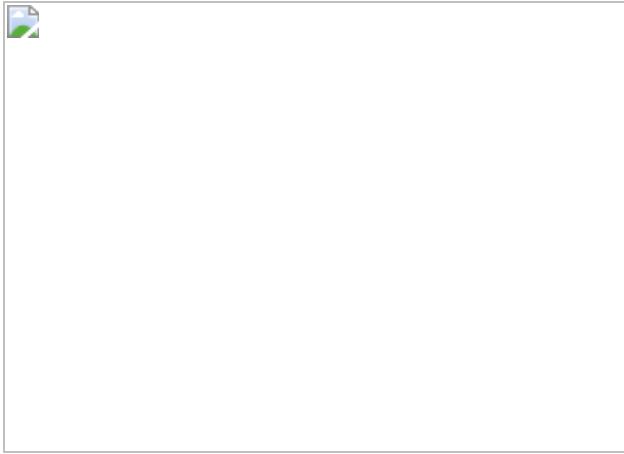
**Fig. 1: Overview of data fusion strategy.**



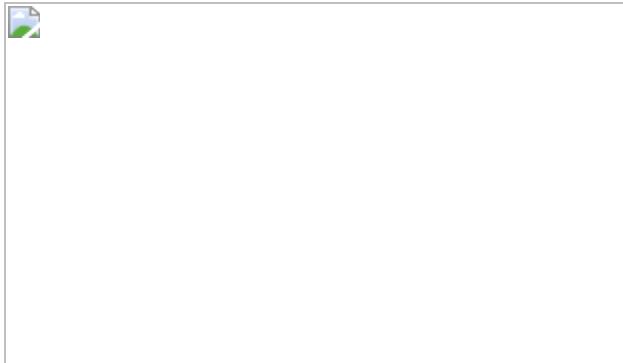
**Fig. 2: The multi-scale integrated cell.**



**Fig. 3: MuSIC captures subcellular components and diameters.**



**Fig. 4: Different data informs different scales of information.**



**Fig. 5: Exploration of MuSIC using physical and functional assays.**



## Data availability

A web portal is available at <http://nrnb.org/music> with links to all major resources used for this study. These include the MuSIC map (<https://doi.org/10.18119/N9188W>); the immunofluorescence (HPA) and AP–MS data (BioPlex 2.0) on which the map is based; and data for the AP–MS pull-down experiments performed as follow-up. The new AP–MS data have also been included as part of the larger compendium of protein interactions in the next version of the BioPlex resource (BioPlex 3.0<sup>29</sup>). AP–MS data, including filtered and unfiltered interaction lists as well as raw mass spectrometry data, are also available at <http://bioplex.hms.harvard.edu>. The image data and associated metadata can also be found in the HPA database (<https://www.proteinatlas.org>). The Gene Expression Omnibus (GEO) accession number for eCLIP data generated in this study is [GSE171553](#). [Source data](#) are provided with this paper.

# Code availability

The MuSIC pipeline is available at <https://github.com/idekerlab/MuSIC> along with a detailed step-by-step guide to building a MuSIC map.

## References

1. 1.

Harold, F. M. Molecules into cells: specifying spatial architecture. *Microbiol. Mol. Biol. Rev.* **69**, 544–564 (2005).

2. 2.

Mori, H. & Cardiff, R. D. Methods of immunohistochemistry and immunofluorescence: converting invisible to visible. In *The Tumor Microenvironment, Methods in Molecular Biology* Vol. 1458 (eds Ursini-Siegel, J. & Beauchemin, N.) 1–12 (Humana Press, 2016).

3. 3.

Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

4. 4.

Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).

5. 5.

Hutlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).

6. 6.

Schaffer, L. V. & Ideker, T. Mapping the multiscale structure of biological systems. *Cell Syst.* **12**, 622–635 (2021).

7. 7.

Ouyang, W. et al. Analysis of the Human Protein Atlas Image Classification competition. *Nat. Methods* **16**, 1254–1261 (2019).

8. 8.

Grover, A. & Leskovec, J. node2vec: scalable feature learning for networks. In KDD ’16: *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 855–864 (2016).

9. 9.

Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep Learning* Vol. 1 (MIT Press, 2016).

10. 10.

Fortunato, S. & Hric, D. Community detection in networks: a user guide. *Phys. Rep.* **659**, 1–44 (2016).

11. 11.

Go, C. D. et al. A proximity-dependent biotinylation map of a human cell. *Nature* **595**, 120–124 (2021)

12. 12.

Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).

13. 13.

Deckert, J. et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* **26**, 5528–5543 (2006).

14. 14.

Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).

15. 15.

Yoshikatsu, Y. et al. NVL2, a nucleolar AAA-ATPase, is associated with the nuclear exosome and is involved in pre-rRNA processing. *Biochem. Biophys. Res. Commun.* **464**, 780–786 (2015).

16. 16.

Chaudhuri, S. et al. Human ribosomal protein L13a is dispensable for canonical ribosome function but indispensable for efficient rRNA methylation. *RNA* **13**, 2224–2237 (2007).

17. 17.

Tafforeau, L. et al. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol. Cell* **51**, 539–551 (2013).

18. 18.

Eppens, N. A. et al. Deletions in the S1 domain of Rrp5p cause processing at a novel site in ITS1 of yeast pre-rRNA that depends on Rex4p. *Nucleic Acids Res.* **30**, 4222–4231 (2002).

19. 19.

De Silva, D., Tu, Y.-T., Amunts, A., Fontanesi, F. & Barrientos, A. Mitochondrial ribosome assembly in health and disease. *Cell Cycle* **14**, 2226–2250 (2015).

20. 20.

Blencowe, B. J. et al. The SRm160/300 splicing coactivator subunits. *RNA* **6**, 111–120 (2000).

21. 21.

The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

22. 22.

Pavan Kumar, P. et al. Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. *Mol. Cell* **22**, 231–243 (2006).

23. 23.

Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).

24. 24.

Fleckner, J., Zhang, M., Valcárcel, J. & Green, M. R. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864–1872 (1997).

25. 25.

Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).

26. 26.

Van Nostrand, E. L. et al. Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). In *mRNA Processing, Methods in*

*Molecular Biology* Vol. 1648 (ed. Shi, Y.) 177–200 (Humana Press, 2017).

27. 27.

Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).

28. 28.

Wang, T. et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* **168**, 890–903 (2017).

29. 29.

Hutlin, E. L. et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022–3040 (2021).

30. 30.

Williams, S. G. & Hall, K. B. Human U2B'' protein binding to snRNA stemloops. *Biophys. Chem.* **159**, 82–89 (2011).

31. 31.

Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. Preprint at <https://arxiv.org/abs/1608.06993> (2016).

32. 32.

Nusinow, D. P. et al. Quantitative proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387–402 (2020).

## Acknowledgements

We thank C. Ng, A. Palmer, Q. Zhang, Y. Quan, members of the laboratories of T.I. and E.L., the Human Protein Atlas and J. Swedlow for discussion and comments; M. Dow for helping us to improve the MuSIC GitHub repository and test the MuSIC pipeline; and the Cell Profiling facility and C. Stadler at the Science for Life Laboratory for help with *in situ* fractionation. This work was supported by the National Institutes of Health (NIH) under grants U54 CA209891, U01 MH115747, P41 GM103504 and R01 HG009979 to T.I., F99 CA264422 to Y.Q., U24 HG006673 to E.L.H., S.P.G. and J.W.H., U41 HG009889 and R01s HL137223 and HG004659 to G.W.Y. and R50 CA243885 to J.F.K.; by a gift from Google Ventures to J.W.H. and S.P.G.; by the Erling-Persson family foundation, Knut and Alice Wallenberg Foundation (2016.0204) and the Swedish Research Council (2017-05327) to E.L.; and by the Belgian Fonds de la Recherche Scientifique (F.R.S./FNRS), the Université Libre de Bruxelles (ULB), the European Joint Programme on Rare Diseases ('RiboEurope' and 'DBAcure'), the Région Wallonne (SPW EER) ('RIBOcancer'), the Internationale Brachet Stiftung and the Epitran COST action (CA16120) to D.L.J.L.

## Author information

### Affiliations

1. Department of Medicine, University of California San Diego, La Jolla, CA, USA

Yue Qin, Marcus R. Kelly, Fan Zheng, Michael Chen, Leah V. Schaffer, Katherine Licon, John J. Lee, Sophie N. Liu, Erica Silva, Jisoo Park, Adriana Pitea, Jason F. Kreisberg & Trey Ideker

2. Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA, USA

Yue Qin, Gene W. Yeo & Trey Ideker

3. Department of Cell Biology, Harvard Medical School, Boston, MA, USA

Edward L. Huttlin, Laura Pontano Vaites, Tian Zhang, Steven P. Gygi & J. Wade Harper

4. Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden  
Casper F. Winsnes, Anna Bäckström, Wei Ouyang & Emma Lundberg
5. Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA  
Maya L. Gosztyla, Steven M. Blue & Gene W. Yeo
6. Stem Cell Program, University of California San Diego, La Jolla, CA, USA  
Maya L. Gosztyla, Steven M. Blue & Gene W. Yeo
7. Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA  
Maya L. Gosztyla, Steven M. Blue, Gene W. Yeo & Trey Ideker
8. RNA Molecular Biology, Fonds de la Recherche Scientifique (F.R.S./FNRS), Université Libre de Bruxelles (ULB), Charleroi-Gosselies, Belgium  
Ludivine Wacheul & Denis L. J. Lafontaine
9. Institute for Artificial Intelligence, Peking University, Beijing, China  
Jianzhu Ma
10. Department of Genetics, Stanford University, Stanford, CA, USA  
Emma Lundberg
11. Chan Zuckerberg Biohub, San Francisco, San Francisco, CA, USA

Emma Lundberg

12. Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

Trey Ideker

13. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA

Trey Ideker

## Contributions

Y.Q., E.L. and T.I. designed the study and developed the conceptual ideas. C.F.W. and W.O. generated image embeddings. Y.Q. and J.M. designed the data integration approach. Y.Q. and F.Z. designed the community detection approach. Y.Q., E.L.H., C.F.W., F.Z., L.V.S., W.O., J.P., A.P., J.F.K., J.M., J.W.H., E.L. and T.I. developed ideas for data analyses. Y.Q. implemented all computational methods and analyses. Y.Q., C.F.W., L.V.S., W.O., J.P. and T.I. organized the GitHub repository and wrote the step-by-step guide. Y.Q., E.L.H., C.F.W., M.R.K., L.P.V., E.S., J.F.K., S.P.G., J.W.H., G.W.Y., D.L.J.L., E.L. and T.I. designed validation experiments. E.L.H., L.P.V., T.Z., J.W.H. and S.P.G. generated and analysed AP–MS data and provided FLAG–HA-tagged clones. S.M.B. and G.W.Y. generated and analysed RIP–qPCR data. L.W. and D.L.J.L. generated and analysed northern blot data. C.F.W., A.B. and E.L. generated and analysed *in situ* fractionation data. M.L.G. and G.W.Y. generated and analysed eCLIP data. Y.Q., M.C., K.L. and J.J.L. performed the rest of the experiments. Y.Q., S.N.L. and T.I. designed the web portal page. Y.Q., E.L. and T.I. wrote the manuscript with input from all authors.

## Corresponding authors

Correspondence to [Emma Lundberg](#) or [Trey Ideker](#).

## Ethics declarations

## Competing interests

T.I. is a co-founder of Data4Cure, is on the Scientific Advisory Board and has an equity interest. T.I. is on the Scientific Advisory Board of Ideaya BioSciences and has an equity interest. G.W.Y is a co-founder, a member of the Board of Directors, on the Scientific Advisory Board, an equity holder and a paid consultant for Locanabio and Eclipse BioInnovations. G.W.Y is a visiting professor at the National University of Singapore. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict-of-interest policies. E.L is on the Scientific Advisory Boards of Cartography Biosciences, Nautilus Biotechnology and Interline Therapeutics, and has an equity interest in all of these. J.W.H. is a co-founder of Caraway Therapeutics, is on the Scientific Advisory Board and has an equity interest. J.W.H. is Founding Scientific Advisor for Interline Therapeutics.

## Additional information

**Peer review information** *Nature* thanks Jason Swedlow and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Characterization of image data used in this study.](#)

**a**, Histogram showing distribution in number of antibodies per protein over 661 proteins included in MuSIC. **b**, Histogram showing distribution in antibody quality scores over antibodies used in this study. **c**, Immunofluorescence images for alternative antibodies (columns) targeting the same protein (rows). Colours represent immunostained protein (green), cytoskeleton (red), or nucleus (blue). Images show high reproducibility for

different antibodies against the same protein. **d**, Comparison of localizations for proteins in MuSIC (HEK293 cells, red) versus all proteins assayed by HPA in any cell line (grey). Localizations as defined by the HPA project<sup>4</sup>.

## Extended Data Fig. 2 Embedding immunofluorescence images and AP–MS data.

**a**, Embedding immunofluorescence (IF) images using DenseNet. The 1024-dimension feature vector for each IF image was extracted from a DenseNet-121<sup>31</sup> model trained to classify the IF image into one or several of 28 pre-defined protein localization classes from HPA. **b**, Two-dimensional visualization (UMAP, n\_neighbours = 5) for the 1,451 image embeddings associated with the 661 proteins in MuSIC. **c**, Ability of different image embedding methods (coloured curves) to generate image-image similarities (cosine similarity) in agreement with protein-protein interactions in BioPlex 2.0. **d**, Node2vec<sup>8</sup> workflow. The feature vector generated by node2vec captures the pattern of interaction neighbourhood for the respective node in input network. **e**, Embedding AP–MS data using node2vec. The input network to node2vec was constructed by treating each protein as a node and assigning edges between protein pairs that were identified as physically interacting in the AP–MS data. The two-dimensional visualization (UMAP, n\_neighbours = 5) for AP–MS embeddings associated with 661 proteins in MuSIC is shown at right. **f**, Network showing all proteins (grey) that physically interact with SNRPC and SNRPB2 (blue) in BioPlex 2.0. SNRPC and SNRPB2 do not physically interact, but the cosine similarity of their embedded features is 0.93 due to shared interaction neighbourhood. In many cases of two proteins with high node2vec similarity but without direct interaction in AP–MS data, we found that neither protein had yet been tagged as bait for an affinity purification experiment. In these cases, the node2vec embedding suggests gaps in existing AP–MS data. **g**, Ability of different AP–MS embedding methods to generate protein-protein similarities (cosine similarity) in agreement with protein pairwise similarities computed from HPA images.

## Extended Data Fig. 3 Fusing protein distances from immunofluorescence and affinity purification.

**a, b**, Protein pairs ranked by similarity in AP–MS embedding enrich for the most similar protein pairs in IF (**a**), and vice versa (**b**). **c**, Calibrating physical diameter,  $D$ , of subcellular components against the number of proteins,  $C$ , assigned to the corresponding Gene Ontology (GO) terms. **d**, Supervised model (random forest) estimates physical proximity (nm) of all pairs of proteins from their IF and AP–MS embeddings. **e**, Performance of model in recovering protein-protein distances in GO in five-fold cross validation (red, Pearson's  $r$ ). Equivalent calculation for random feature sets (grey). Statistics calculated using two-sided paired t-test. Data are presented as mean values +/- standard deviation.

## Extended Data Fig. 4 Selection of parameters for community detection.

**a**, Using multi-scale community detection, protein systems of increasing sizes are discovered as the threshold for protein-protein distance is progressively increased. **b**, CliXO community detection has four parameters (depth  $\alpha$ , y-axis; breadth  $\beta$ , x-axis; minimum modularity  $m$  and modularity significance  $z$ , red circle backslash) that affect the sensitivity with which communities are identified and thus the size of the hierarchy. **c, d**, Dot plots in which each dot is a community hierarchy generated with a particular set of parameters. The selection for MuSIC is highlighted in red. This selection was among several that were optimal based on enrichment for protein-protein interactions in Human Cell Map (**c**) and co-essentialities from DepMap (**d**). Examples of other parameter sets are shown in blue. **e**, Map from Fig. 2 with system colour showing enrichment for co-essentialities among protein pairs that are specific to that system. Enrichment of each system is assessed empirically, using 1,000 randomized hierarchies, followed by Benjamini–Hochberg multiple test correction to obtain FDR (orange gradient).

## Extended Data Fig. 5 Supporting analyses for PRRPA.

**a**, Distributions of protein-protein distance z-scores among the seven proteins in the PRRPA system for IF (top, red) or AP–MS (bottom, blue) modalities, calibrated to all such distances, respectively (grey). Statistics calculated using one-sided Mann–Whitney *U* test. **b**, Specific recovery of new AP–MS interactions within PRRPA is shown (dark blue bar), in comparison to interactions between proteins in PRRPA and other proteins organized under the same parent systems (“Ribosome” and “Ribosome biogenesis assembly”, light blue bar), or between proteins in PRRPA and those organized elsewhere in MuSIC (grey bar). **c**, Mature 28S/18S rRNA ratio under siRNAs targeting each PRRPA protein (green) versus scrambled siRNA (grey),  $n = 3$  biological replicates. FDR from two-sided t-test with Benjamini–Hochberg correction. Data are presented as mean values +/- standard deviation. **d–i**, Western blot analysis (**d**, **e**, Simple western assay; **f–i**, SDS–PAGE) of target protein abundance after treating HEK293T cells with respective siRNA for 72 h (Supplementary Tables [6](#), [7](#)). The siRNAs highlighted in red were selected to assess the perturbation of mature rRNA ratio (28S/18S rRNA) when knocking down target protein, with protein knockdown efficiency confirmed using western blot in three additional biological replicates. For source data, see Supplementary Fig. [1](#) (gel; **d–i**) and Supplementary Fig. [2](#) (total RNA profiles; **c**).

## Extended Data Fig. 6 Supporting analyses for ribosomal systems.

**a**, Categorization of proteins in “Ribosome biogenesis community” by whether they have been previously identified in human ribosome biogenesis. Excludes PRRPA proteins described in Fig. [5b–d](#). **b**, Structure of human pre-rRNA and probes used for northern blot. In eukaryotes, 3 out of 4 mature rRNAs (18S, 5.8S, and 28S rRNAs) are produced from a single long polycistronic precursor (47S) synthesized by RNA polymerase I. The mature rRNAs are interspersed with the 5' and 3' external transcribed spacers (ETS) and internal transcribed spacer (ITS) 1 and 2. The probes used in the northern blot (5'-ETS, ITS1, and ITS2) are indicated and colour-coded. **c**, Total RNA extracted from the indicated cell line, which was transfected with a DsiRNA specific to the target protein for 72 h and analysed by northern blotting with probes specific to the 5'-ETS, ITS1, and ITS2 sequences (Supplementary Table [8](#)). As controls, cells were either

untreated, transfected with a scrambled silencer, or transfected with a silencer targeting UTP18 (positive control involved in small ribosomal subunit biogenesis). Heat map colour shows the percentage of each pre-rRNA species with respect to the scramble control. For gel source data, see Supplementary Fig. 1. **d**, For protein baits in new AP–MS experiments (x axis), fraction of interacting preys that fall within the Ribosome biogenesis community (blue bars) versus elsewhere (grey bars). Only new AP–MS interactions are considered for this analysis. RNPS1 does not belong to Ribosome biogenesis community and serves as a negative control. **e**, IF images showing similar cytoplasmic staining for proteins in “Mito-cyto ribosomal cluster.” Cytoplasmic staining is dim for MRPS9, MRPS14 and MRPS31 compared to their predominant mitochondrial locations. Colours represent immunostained protein (green), cytoskeleton (red) and nucleus (blue). **f, g**, Corresponding distributions of protein-protein distance z-scores for IF (**f**, red) or AP–MS (**g**, blue), calibrated to all such distances, respectively (grey). Statistics calculated using one-sided Mann–Whitney *U* test. **h**, Two-dimensional projection of proteins in Mito-cyto ribosomal cluster, as in Fig. 5f. Proteins coloured according to known affiliations to cytoplasmic ribosome or mitochondrial ribosome. **i**, Validated AP–MS interactions in Mito-cyto ribosomal cluster. Note that only one out of seven proteins was previously tagged as bait in BioPlex 2.0 (light blue node), thus most physical associations (dark blue edges) among protein pairs were newly identified in this study.

### **Extended Data Fig. 7 Supporting analyses for chromatin regulation and splicing systems.**

**a**, IF images showing similar nucleoplasm and nuclear speckles signals among proteins in the “Chromatin regulation complex.” Colours represent immunostained protein (green) and cytoskeleton (red). **b**, Distributions of pairwise protein distance z-scores among the proteins in the Chromatin regulation complex for IF (top, red) or AP–MS (bottom, blue) modalities, calibrated to all such distances, respectively (grey). Statistics calculated using one-sided Mann–Whitney *U* test. **c**, Immunofluorescent proteins (rows) imaged in HEK293 cells, untreated (left) or treated (right) with *in situ* fractionation to remove soluble cytoplasmic and loosely held nuclear proteins. Chromatin-binding proteins remain after treatment. Blue, nucleus;

other colours as in **a**. For image source data, see Supplementary Fig. [3](#). **d**, IF images showing similar nucleoplasm signals among proteins in “RNA splicing complex 3.” **e**, Similar display for RNA splicing complex 3 as in **b**. **f**, Comparison of 500 top differentially expressed mRNAs (absolute fold change) resulting from shRNA knockdown of each of five genes (see Supplementary Table [9](#) for file accessions). Bar chart shows number of differential mRNAs shared by different gene groups indicated by black dots beneath each bar. One-sided one-sample t-test. **g**, Comparison among the top 10 pathways (Gene Ontology Biological Process) returned from Gene Set Enrichment Analysis using the top 500 differentially expressed transcripts. Bar chart shows number of enriched pathways shared by different gene groups indicated by black dots beneath each bar. One-sided one-sample t-test. **h**, eCLIP workflow. RBP, RNA-binding protein. NGS, next generation sequencing.

## Extended Data Fig. 8 Supporting analyses for Discussion.

**a, b**, Examples of proteins with strong AP–MS protein interactions that have very different IF localization patterns. Colours represent immunostained protein (green) and cytoskeleton (red). **c**, Degree of co-essentiality for gene pairs within PRRPA (teal bar) shown in comparison to remaining pairs of genes assigned to the more general system that contains it, “Ribosome biogenesis community” (green bar), as well as all other gene pairs in MuSIC (grey bar). **d**, Similar analysis as in **(c)** for “RNA splicing complex 3.” Parent systems are “RNA processing complex 1” and “RNA splicing complex family.” **e**, Protein co-abundance for MuSIC systems, calculated from the median Pearson correlation of pairwise protein abundance over 375 diverse cell lines<sup>[32](#)</sup>. The plot shows all systems with fewer than 20 proteins and co-abundance measurements for >50% of protein pairs. Significance is assessed empirically (one-sided), using 1,000 randomized MuSIC hierarchies, followed by Benjamini–Hochberg multiple test correction to obtain FDR (colour of bar). Protein co-abundance for a system provides evidence for its presence in cell types beyond HEK293.

## Supplementary information

## **Supplementary Figures**

This file is consisted of three supplementary figures and provides the source gel data (Supplementary Figure 1), total RNA profiles (Supplementary Figure 2) and source in situ fractionation data (Supplementary Figure 3).

## **Reporting Summary**

## **Supplementary Methods**

This file contains Supplementary Methods and Supplementary References.

## **Peer Review File**

## **Supplementary Table 1**

MuSIC proteins and associated data.

## **Supplementary Table 2**

Literature collection of subcellular components used for calibrating physical diameter (related to Extended Data Fig. 3c).

## **Supplementary Table 3**

MuSIC systems and associated data.

## **Supplementary Table 4**

Literature collection of subcellular components used for validating MuSIC estimated diameter (related to Fig. 3b).

## **Supplementary Table 5**

866 reproducible and significant (IDR cut-off of 0.01, Fisher's Exact test  $P \leq 0.001$ , fold enrichment  $\geq 8$ ) eCLIP peaks of RPS3A.

## **Supplementary Table 6**

Sequences of siRNA and DsiRNA used in this study.

## **Supplementary Table 7**

Antibodies used in this study.

## **Supplementary Table 8**

Sequences of northern blot probes used for pre-rRNA analysis (related to Fig. 5e and Extended Data Fig. 6b, c).

## **Supplementary Table 9**

ENCODE file accessions used for RNA-seq analysis (related to Extended Data Fig. 7f, g).

## **Source data**

### **Source Data Fig. 5**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Qin, Y., Huttlin, E.L., Winsnes, C.F. *et al.* A multi-scale map of cell structure fusing protein images and interactions. *Nature* **600**, 536–542 (2021). <https://doi.org/10.1038/s41586-021-04115-9>

- Received: 18 June 2020

- Accepted: 08 October 2021
- Published: 24 November 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04115-9>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04115-9>

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 01 December 2021](#)

# Accuracy mechanism of eukaryotic ribosome translocation

- [Muminjon Djumagulov](#) [ORCID: orcid.org/0000-0002-6900-0768](#)<sup>1,2</sup>,
- [Natalia Demeshkina](#)<sup>1</sup> [nAff3](#),
- [Lasse Jenner](#) [ORCID: orcid.org/0000-0003-3917-8853](#)<sup>1</sup>,
- [Alexey Rozov](#) [ORCID: orcid.org/0000-0003-2545-1978](#)<sup>1,2</sup>,
- [Marat Yusupov](#) [ORCID: orcid.org/0000-0001-5544-0597](#)<sup>1</sup> &
- [Gulnara Yusupova](#) [ORCID: orcid.org/0000-0002-9119-9136](#)<sup>1</sup>

*Nature* volume 600, pages 543–546 (2021)

- 5475 Accesses
- 27 Altmetric
- [Metrics details](#)

## Subjects

- [Ribosome](#)
- [X-ray crystallography](#)

## Abstract

Translation of the genetic code into proteins is realized through repetitions of synchronous translocation of messenger RNA (mRNA) and transfer RNAs (tRNA) through the ribosome. In eukaryotes translocation is ensured by elongation factor 2 (eEF2), which catalyses the process and actively contributes to its accuracy<sup>1</sup>. Although numerous studies point to critical roles for both the conserved eukaryotic posttranslational modification diphthamide in eEF2 and tRNA modifications in supporting the accuracy of translocation, detailed molecular mechanisms describing their specific functions are poorly understood. Here we report a high-resolution X-ray

structure of the eukaryotic 80S ribosome in a translocation-intermediate state containing mRNA, naturally modified eEF2 and tRNAs. The crystal structure reveals a network of stabilization of codon–anticodon interactions involving diphthamide<sup>1</sup> and the hypermodified nucleoside wybutosine at position 37 of phenylalanine tRNA, which is also known to enhance translation accuracy<sup>2</sup>. The model demonstrates how the decoding centre releases a codon–anticodon duplex, allowing its movement on the ribosome, and emphasizes the function of eEF2 as a ‘pawl’ defining the directionality of translocation<sup>3</sup>. This model suggests how eukaryote-specific elements of the 80S ribosome, eEF2 and tRNAs undergo large-scale molecular reorganizations to ensure maintenance of the mRNA reading frame during the complex process of translocation.

[Download PDF](#)

## Main

In eukaryotes, the complex process of translocation is ensured by eEF2, a GTPase that is indispensable for maintaining the correct mRNA reading frame. Many genetic and biochemical studies point to a critical role of the unique eEF2 post-translational modification diphthamide, which is located in domain IV and is conserved among eukaryotes and archaea. Organisms lacking diphthamide have reduced protein synthesis rates and increased occurrence of (−1) frameshifting<sup>1,4</sup>. Diphthamide is a target of several virulent toxins that inactivate eEF2 by ADP ribosylation and cause lethal effects<sup>5</sup>.

The current structural knowledge about dynamics of the eukaryotic elongation cycle has been provided by low-to-intermediate resolution cryo-electron microscopy (cryo-EM) reconstructions<sup>6,7,8,9,10</sup>. Although several late steps of translocation have been described, the level of detail achieved in these studies is insufficient to suggest a precise mechanism explaining fidelity of translocation. At the same time, understanding of the principles of the fidelity has become particularly crucial during the current viral pandemic, because programmed mRNA frameshifting is at the heart of the SARS-CoV-2 replication cycle<sup>11</sup>.

Here we present the crystal structure of eukaryotic 80S ribosomes from *Saccharomyces cerevisiae* trapped in intermediate translocation state. The X-ray crystal structure provides a detailed mechanism of tRNA translocation from A- to P-sites and highlights the specific role of eEF2 in the movement of the tRNA–mRNA module during the process. The crystal structure presented here uncovers the precise role of diphthamide and wybutosine, a heavily modified nucleoside at position 37 of eukaryotic phenylalanine tRNA, in stabilization of the codon–anticodon interactions during translocation and demonstrates how eukaryote-specific elements of the 80S

ribosome, eEF2 and tRNA rearrange to ensure maintenance of the mRNA reading frame.

## Architecture of the translocation complex

We determined the structure of the *S. cerevisiae* 80S ribosome translocation complex trapped in intermediate state by X-ray crystallography at 3.2 Å resolution (Fig. [1a](#), [b](#), Extended Data Fig. 1, Extended Data Table [1](#)). It consists of *S. cerevisiae* 80S ribosomes bound with native *S. cerevisiae* eEF2, the nonhydrolyzable GTP analogue GMPPCP, mRNA and two tRNAs, and was determined in the absence of antibiotics, which customarily used for stabilization, suggesting that the model represents a bona fide state of the 80S ribosome.

**Fig. 1: The translocation-intermediate state of the eukaryotic 80S ribosome with eEF2–GMPPCP, mRNA and tRNAs,, showing diphthamide of eEF2 is involved in stabilizing codon–anticodon interactions early in translocation.**

---

 **figure1**

**a**, Overview of the translocation-intermediate complex with two tRNAs trapped in chimeric hybrid ap/P and pe/E states. **b**, Close-up view of ap/P and pe/E tRNA anticodon stem loops in the context of elements of the SSU body and head and domain IV of eEF2. Position of wybutosine (yW) of ap/P tRNA<sup>Phe</sup> is indicated by the asterisk. **c**, Position of diphthamide (Diph699) at the conjunction of the AAG anticodon of ap/P tRNA<sup>Phe</sup>, mRNA codon UUC and decoding adenosines 1755–1756 of helix 44 (h44). **d**, Stabilization networks around codon–anticodon interactions in the translocation-intermediate complex (left) and the decoding centre of the bacterial ribosome in the cognate classical state<sup>13,14</sup> (right). Top, middle and bottom panels depict stabilization around the first, second and third base pairs (BP1–3) of a codon–anticodon duplex, respectively. Conserved adenosines 1755 and 1756 in yeast 18S rRNA correspond to

adenosines 1492 and 1493 in 16S rRNA of the bacterial decoding centre. eEF2 is shown in red, mRNA is in orange, chimeric hybrid ap/P tRNA is in green and chimeric hybrid pe/E is in yellow. LSU rRNA and proteins are shown in grey and purple; SSU rRNA and proteins are in cyan and in deep blue. Degrees of SSU body and head rotations are indicated and were obtained by superimposing with the non-rotated 80S ribosome (PDB ID 3J78).

The crystal structure of the ribosome complex represents an intermediate translocation state that has not been described before with two tRNAs trapped in the chimeric hybrid ap/P and pe/E transitory positions. In this state the small subunit (SSU) head has swivelled 13.5° and the SSU body has undergone 9° anticlockwise rotation relative to the large subunit (LSU) (Fig. 1a). The anticodon stem-loop (ASL) of the A-site tRNA is captured half-translocated between the A- and P-sites of SSU (12.4 Å out of a fully translocated distance of 24.1 Å), and the tRNA acceptor end contacts the P-loop of the peptidyl-transferase centre of LSU forming an ap/P chimeric hybrid state (Fig. 1a,b, Extended Data Fig. 2a,b).

The overall conformation of eEF2 resembles that of eEF2 in the 80S translocation intermediate–post-translocation (TI-POST)-1 state of a medium-resolution cryo-EM reconstruction<sup>6</sup> that depicts a late-translocation event with domain IV of eEF2 protruding approximately 7 Å deeper into the A-site of SSU from the perspective of the SSU body (Extended Data Fig. 3). However—in contrast to the TI-POST-1 state—in our complex, domain IV of eEF2 is engaged in an extensive stabilization network with the codon–anticodon duplex and with the decoding centre area (Fig. 1b–d, Extended Data Fig. 1d). The intricate network of arginine-rich regions of domain II and III of eEF2 with helix 5 of 18S rRNA is formed owing to strong rotation of the SSU body (9.85°). These contacts are not present in the less rotated (4.26°) TI-POST-1 or unrotated TI-POST-3 states of 80S from the cryo-EM models<sup>6</sup> (Extended Data Fig. 4a,b). This comparison suggests that eEF2 uncouples from the SSU body at the later steps of translocation where it is associated only with LSU and remains stationary relative to the rest of the ribosome. Collectively, these structural data corroborate an existing hypothesis describing a translocase (eEF2 or EF-G) as a ‘locking pawl’ that decouples tRNA–mRNA from the SSU head and body during back-rotation<sup>3</sup>.

Comparison of the current early-intermediate state with late-translocation states described by cryo-EM<sup>6</sup> reveals that the uS12 protein remains attached to domain III of eEF2 during SSU back-ratcheting. Located on the SSU shoulder, uS12 is implicated in translocation<sup>12</sup> and codon–anticodon duplex stabilization in the decoding centre<sup>13</sup>. It is plausible that during translocation, uS12 transmits SSU body back-rotation by pulling domain III of eEF2 that in turn retransmits it to switch II in the G-domain, stimulating GTP hydrolysis (Extended Data Fig. 4c,d). We found additional stabilization of domain IV by the N terminus of the eukaryote-specific protein eS30, which also

interacts extensively with conserved decoding protein uS12 (Extended Data Fig. 4c). It can be assumed that eS30 co-evolved with eEF2, whose domain IV has 65 additional amino acids compared with its bacterial counterpart EF-G, to provide supplementary stabilization as well as enhancing propagation of conformational changes at the decoding site<sup>13,14</sup>. Other interactions between eEF2, tRNAs and the 80S ribosome are presented in Extended Data Figs. 5, 6.

## Diphthamide

The current crystal structure of 80S ribosome translocation intermediate reveals direct involvement of diphthamide in stabilization of a codon–anticodon duplex in transition from the A- to P-site. eEF2 interacts with the mRNA codon exclusively via diphthamide, which protrudes into a cleft formed by mRNA, ap/P tRNA and rRNA (Fig. 1c, d). Assisted by His583 and Asp696 of eEF2, diphthamide contacts the codon–anticodon duplex minor groove. A similar pattern of interactions has been extensively described between the bacterial decoding centre mould and the codon–anticodon duplex in a classical unrotated state<sup>13,14</sup> (Fig. 1d). The closest resemblance is observed at the second base pair (BP2, Fig. 1d), where diphthamide together with Asp696 mimics stabilization of decoding nucleotides G577 and A1755 in 18S rRNA (bacterial G530 and A1492). Fixation of the first codon–anticodon pair is divided between diphthamide with His583 and His694 contacting the anticodon ribose, and the wybutosine modification of ap/P tRNA<sup>Phe</sup> of nucleotide 37 stacking on codon position +4 (BP1, Figs. 1d, 2a). The third codon–anticodon pair (BP3, Fig. 1d) is anchored by diphthamide only through its trimethylammonio moiety stabilizing the codon position. These findings are corroborated by numerous studies describing increased –1 frameshifting slippage when translation is performed with eEF2 lacking diphthamide<sup>1,15</sup>.

**Fig. 2: Stabilization of mRNA–tRNA by wybutosine modification of tRNA<sup>Phe</sup>, rearrangements of the decoding centre and depiction of a pawl function for eEF2 during translocation.**

---

 **figure2**

---

**a**, Wybutosine cross-strand stacks on the first codon–anticodon base pair formed by mRNA codon UUC and tRNA<sup>Phe</sup> at the ap/P state and stabilizes the last position of the adjacent upstream AUG codon (G3). An alternative view (bottom) of the same pattern clearly demonstrates how wybutosine enhances stabilization of mRNA to prevent a frame shift during translocation. **b**, Interactions between the backbones of translocating mRNA and decoding A1756 and A1757 of h44; mRNA in classical state is in grey. **c**, A shift of helix 69 (H69) of 25S rRNA is coupled to rearrangements of the decoding centre as compared to the bacterial classical state in white (PDB ID 4V6F). **d**, Contacts of eEF2 (coloured by domain) with LSU (grey), body (cyan) and head (pale cyan) of SSU; dashed squares indicate contacts that are disrupted at the late stage of translocation, allowing eEF2 to uncouple from the SSU body. **e**, Superposition of intermediate- and late (grey) (TI-POST-1, PDB ID 6GZ3)-translocation structures relative to LSU reveals eEF2 acting as a pawl anchoring and decoupling ap/P tRNA from the SSU head and body. The reverse direction of the SSU head and body rotation to a classical post-translocation state is indicated with arrows.

The mRNA backbone of the UUC codon paired to ap/P tRNA<sup>Phe</sup> is located in close proximity to the sugar–phosphate backbone of decoding nucleotides A1755-A1756 and G1757, whose backbone is bulged out because of partial ‘flipped-in’ and ‘flipped-out’ positions of their nucleoside moieties (Fig. 2b). This close positioning of mRNA and A1756-G1757 backbones is realized by ribose–phosphate zipper interactions<sup>16</sup>. The diphthamide trimethylammonio moiety additionally stabilizes negatively charged backbones by interaction with the phosphate group between A1756 and G1757 of 18S rRNA (BP3, Fig. 1d).

The difference in structure of helix 44 of SSU rRNA (h44), which forms a part of the decoding centre between bacteria and both eukaryotes and archaea might explain emergence of the diphthamide modification during evolution (Extended Data Fig. 7).

The changes of decoding centre bulge, and flanking nucleotides could lead to increased flexibility of the eukaryotic decoding centre, which would require additional stabilization that was achieved by development of diphthamide modification of eEF2. In addition, the unique trimethylammonio moiety could have been evolutionally refined to reduce repulsion between juxtaposing negatively charged backbones of mRNA and the h44 decoding-centre loop at the early stages of translocation.

## Unlocking of the decoding centre

In the early-intermediate translocation state, when diphthamide takes over stabilization of the codon–anticodon duplex, decoding nucleotides rearrange to initiate resetting of the decoding centre. Nucleotide A1755 of 18S rRNA adopts a flipped-in position, whereas A1756 is in a more flipped-out and somewhat flexible conformation. There is a noticeable rearrangement of the pivotal intersubunit bridge B2a that is partially composed of A2256 of LSU helix 69 (H69) and decoding nucleotides of h44 (A1755–A1756) of SSU (Fig. [2c](#)). In the present structure, adenine 2256 of H69 protrudes towards sugar phosphate backbone between A1755–A1756, and contacts A1756 phosphate and A1755 ribose moieties. This connection allows H69 to transmit its movement directly to decoding adenines.

During selection of tRNA on the ribosome, the decoding centre serves as a mould imposing restraints on codon–anticodon nucleotides via defined interactions<sup>[13,14](#)</sup> (Fig. [1d](#), right). To translocate mRNA by one codon, the decoding centre has to unlock from the bound codon–anticodon duplex. The present crystal structure with eEF2 suggests that destabilization of the decoding centre is initiated by anticlockwise rotation of SSU, which leads to rearrangement of the B2a bridge. The latter rearrangement induces displacement of H69 towards decoding h44, resulting in a new contact of A2256 of H69 with the decoding nucleotides A1755–A1756. Further rearrangements of the H69 tip lead to its synchronized movement with the decoding adenosines and triggers partial unlocking of the decoding centre from mRNA and tRNA (Fig. [2c](#)). The trimethylammonio moiety of diphthamide contributes to these changes of A1755–A1756 and prevents decoding adenines from re-establishing their contacts with the codon–anticodon duplex (Fig. [1c,d](#)).

Diphthamide-induced unlocking catalyses translocation by reducing the energy required for movement of the codon–anticodon pair. This function is supported by biochemical studies, which have demonstrated decreased protein synthesis rates in organisms lacking diphthamide<sup>[4,15](#)</sup>. Such an interpretation is also consistent with results of pre-steady-state kinetics studies on the bacterial translocation system<sup>[17](#)</sup> and may explain the inhibitory effects of the antibiotics paromomycin and viomycin that interfere with the resetting of the decoding nucleotides by binding to the decoding

centre region of h44 and reduce the rate of translocation by about 160- and more than 10,000-fold, respectively<sup>18,19,20</sup>.

## Wybutosine

In the current structure, we observe clear density for the tRNA<sup>Phe</sup> hyper-modification wybutosine in position 37 in an authentic binding state on the ribosome. Wybutosine consists of wyosine base with 4-methoxy-3-[ (methoxycarbonyl)amino]-4-oxobutyl group (group R; Fig. 2a, Extended Data Fig. 2c). Wyosine cross-strand stacks with the first codon nucleoside paired to ap/P tRNA. Group R stretches towards the third nucleotide (G3) of the codon coupled to the anticodon of pe/E tRNA and forms two hydrogen bonds between its methylcarboxyl portion and G3, hence, upgrading the triplet codon–anticodon interaction with ap/P tRNA to a quadruplet interaction (Fig. 2a). To our knowledge, this is the first time a tRNA modification has been seen to directly influence two adjoint codon–anticodon pairs by interactions with mRNA. Interactions between group R of wybutosine and the third position of the adjacent codon coupled to pe/E tRNA are possible because of a closer distance between ap/P tRNA and pe/E tRNA (Extended Data Fig. 2a). This situation is not achievable with tRNAs in classical A/A and P/P states and during the early stage of translocation where both tRNAs are positioned more than 10 Å apart, indicating that the main stabilization role of wybutosine modification is strongly manifested in the intermediate translocation states. This also implies that the (−1) ribosomal frameshifting events with hypomodified wybutosine tRNA are likely to occur during intermediate and late steps of translocation. These findings substantiate previous studies explaining why lack or alteration of wybutosine derivatives intRNA<sup>Phe</sup> increases the (−1) programmed ribosomal frameshifting frequency and is associated with poor survival in patients with cancer<sup>21</sup>, and also demonstrates how the presence of wybutosine influences the efficiency of viral ribosomal frameshifting<sup>22</sup>.

## Conclusion

The unique state of the translocating eukaryotic ribosome described here demonstrates eEF2 functioning as a pawl<sup>3</sup> during translocation (Figs. 2d, 3, Extended Data Fig. 4a, b). Comparison of the early translocation-intermediate complex described here with the late translocation-intermediate TI-POST-1<sup>6</sup> shows that eEF2 is sturdily anchored on LSU via domains I and V (Fig. 2e, [Supplementary Video](#)) with tRNAs retaining very similar positions in both states of the 80S ribosome, while SSU head and body move around them. The comparison suggests that domain IV of eEF2 uncouples tRNA–mRNA from the SSU body and head allowing these domains to return to their pretranslocation positions without pulling tRNA and mRNA back with them. Interestingly, an increase in the head swivel from around 13.5° (in the early

intermediate) to about 18° (in TI-POST-1) is a result of the body back-rotation that the head cannot follow because it remains attached to tRNAs (Fig. 2e). This probably creates a strain in the SSU neck region that eventually forces the head to unbind from tRNAs and to rotate back ([Supplementary Video](#)). Thus, the last steps of translocation are achieved by back retraction of the head to a non-rotated state and rebinding of tRNA–mRNA (one codon further down) to the SSU P-site environment.

**Fig. 3: Integrating kinetic and structural studies of translocation.**



Top, translocation scheme based on the crystal structure of the intermediate translocation complex reported here (in frame) and on cryo-EM structures of late translocation (PDB ID: 6GZ3 and 6GZ5), as well as hybrid and classical post-translocation states (PDB ID: 3J77 and 3J78). A proposed sequence of events based on kinetic studies<sup>28</sup> is shown at the bottom. Steps 1 and 2, thermally driven intersubunit rotations lead to tRNAs adopting hybrid A/P and P/E states and eEF2–GTP binding to the 80S ribosome. Steps 2 and 3, concomitant changes of LSU H69 composing intersubunit bridge B2a and the decoding centre, and insertion of the eEF2 diphthamide to the SSU A-site induce unlocking of the decoding centre. The released codon–anticodon duplex becomes stabilized by direct interactions with diphthamide. Detachment of tRNA ASLs from the SSU body and further insertion of the eEF2 domain IV into the A-site cause initial anticlockwise rotation of the head and movement of the second tRNA from the SSU P-site towards the E-site where it binds to L1 stalk. Steps 3 and 4, eEF2 remains anchored to LSU via domains I and V but is released from SSU where domain IV uncouples tRNA–mRNA from rearrangements of the SSU body and head. What is perceived as an additional large swivelling of the head is actually a result of the body back-rotation while the head remains fixed to tRNAs. Step 4 and 5, this body rotation increases the strain in the SSU neck and leads to uncoupling of the head from tRNAs. Formation of contacts between rRNA of the head and domain IV of eEF2 restrain the head position. The last steps of translocation

are achieved when the head, owing to the increasing strain on the neck, snaps back to a non-rotated state and tRNA–mRNA binds to the SSU P-site environment.

The translocation process is enabled by Brownian intersubunit rotations of the ribosome<sup>23,24</sup>. The GTP-bound state of eEF2 on the ribosome stays in rigid conformation, as it is observed in current structure of early-translocation intermediate as well as in late-translocation intermediates<sup>6</sup> (Fig. 2d,e), thus serving as a pawl that ensures directionality of translocation process. The described model supports findings showing that the energy stored in the eEF2•GTP state is sufficient to promote translocation<sup>25</sup> and suggests that hydrolysis of GTP does not occur until the late steps of the process. It has been shown previously that domain IV of eEF2•GTP state can reach a codon–anticodon duplex of the P-site tRNA of the non-rotated ribosome<sup>6</sup>. However, experiments using fluorescence resonance energy transfer and other techniques<sup>6,26</sup> have reported that translocation induced by eEF2 with a non-hydrolyzable analogue of GTP is prone to reversion, demonstrating a critical role of GTP hydrolysis in promoting unidirectionality of translocation. GTP hydrolysis and inorganic phosphate release are most likely to occur during the late steps of translocation during SSU body back-ratcheting, when hydrolysis of GTP is stimulated by the movement of uS12–eEF2 domain III (Extended Data Fig. 4c,d). Conformational changes of eEF2 induced by GTP hydrolysis enable unbinding from the mRNA–tRNA module in a manner that prevents pulling the codon–anticodon duplex back to the A-site. Similarly, the bacterial homologue of eEF2 undergoes a large rotation in domain III before its dissociation from the ribosome<sup>27</sup>.

## Methods

### 80S ribosome purification

Purification of the 80S ribosomes from the JD1370-ΔStm1 yeast strain<sup>7</sup> was carried out according to the previously described protocol<sup>29</sup>, with some modifications. The crude ribosomes obtained by precipitation with 8.5% PEG 20,000 were re-suspended in buffer M (30 mM Hepes-KOH, 10 mM MgCl<sub>2</sub>, 50 mM KCl, 8.5% mannitol, 0.5 mM EDTA-KOH, 2 mM DTT, pH 7.55) and MgCl<sub>2</sub> and KCl concentrations were slowly adjusted to 10 and 500 mM (10/500), respectively. The ribosomal suspension was then incubated on ice for 35 min with mild vortexing. The ribosomes were applied on the 6% sucrose cushion, which was prepared in the same dissociation conditions (10/500) and layered over the 10–30% sucrose gradient as in ref. <sup>7</sup> with 5 mM spermidine (5 mg of ribosomes per SW28 tube). After the overnight centrifugation selected fractions of the 80S ribosomes were collected and the ribosomes were precipitated by PEG 20,000.

## Purification of native eEF2

The isolation procedure of native eEF2 was mainly based on the previously described protocol<sup>30</sup>, with changes in several steps. First, we used a fresh culture of yeast strain JD1370-ΔStm1 grown to an  $A_{600}$  of 5–6 and cells were lysed in a microfluidizer. Second, instead of S-Sepharose, source-Q and uno-Q ion-exchange columns we used SP-Sepharose, Q-Sepharose and introduced a gel filtration with Sephadex-200 as the final purification step. The final sample was stored in pH 7.5 buffer consisting of 20 mM Tris-HCl, 5 mM MgCl<sub>2</sub>, 100 mM KCl, 10% glycerol and 1 mM DTT.

## Purification and aminoacylation of tRNAs

*S. cerevisiae* tRNA<sup>Phe</sup> ('chemical block') was aminoacylated according to the protocol as described<sup>31</sup> with minor modifications. After three rounds of phenol–chloroform–isoamyl alcohol extraction *S. cerevisiae* Phe-tRNA<sup>Phe</sup> was purified on a column DeltaPack, C4 300A, 5mm, 3.9 × 150 mm HPLC column (Waters) using a ethanol gradient as described<sup>32</sup>. The final sample was stored in 20 mM NH<sub>4</sub>CH<sub>3</sub>CO<sub>2</sub> pH 5.0 at –80 °C. *Escherichia coli* tRNA<sup>fMet</sup> was prepared and then aminoacylated and formylated according to<sup>33</sup>. After phenol extraction, fMet-tRNA<sup>fMet</sup> was purified by hydrophobic chromatography using TSK-gel Phenyl-5PW column, and the final sample was stored in 20 mM NH<sub>4</sub>CH<sub>3</sub>CO<sub>2</sub> pH 5.0 at –80 °C.

## 80S ribosome translocation complex formation

For reconstitution of translocation complex *S. cerevisiae* 80S ribosomes (2.2 μM) and 5'-AAUGUUCAA-3' mRNA (Dharmacon) (2.9 μM) were incubated at 30 °C for 10 min in 6 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 50 mM KOAc, 10 mM NH<sub>4</sub>Cl and 1.25mM DTT, 10 mM Hepes-KOH (pH 7.5). The fMet-tRNA<sup>fMet</sup> (2.9 μM) was added and the complex further incubated for 7 min at 30 °C with following addition of Phe-tRNA<sup>Phe</sup> (6.5 μM). The complex was incubated for additional 7 min at 30 °C. Separately, *S. cerevisiae* eEF2 (8.7 μM) was incubated with GDPCP (Jena Bioscience) (0.25 mM) for 10 min at room temperature and mixed with the ribosome complex for a final incubation at 30 °C for 10 min. The detergent Deoxy Big CHAP (CalBioChem) (2.4 mM) was added and after 5 min at room temperature the complex was incubated at 4 °C for 5 min.

## Crystallization and crystal treatment

The 80S ribosome translocation complex was crystallized at 4 °C by vapour diffusion in the MRC-48 sitting drop plates (Hampton Research) by mixing 3 μl of the complex with 3 μl of the reservoir solution (100 mM bis-Tris-HCl, pH 7.0, 300 mM NH<sub>4</sub>SCN,

100 mM KCl, 8.25% – 9.5% PEG 20K, 1 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 2% glycerol, 1% sucrose, 5 mM putrescine). Crystals appeared after 3 days and grew to their full size in 13 days.

The post-crystallization treatment was carried out via dehydration by replacing reservoir solution with saturated MgCl<sub>2</sub> salt and incubating for 18 h. Treatment solution (3.3% PEG 20K, 6% PEG 10K, 115 mM bis-Tris-HCl, pH 5.4, 18 mM putrescine, 21 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 9% glycerol, 0.75% sucrose, 1.8 mM deoxy big CHAP, 2.3 mM DTT) was added to the crystallization drop before dehydration. The crystals were collected and flash-frozen in liquid nitrogen.

## Data collection and structure refinement

Data collection was carried out at 100 K at beamline PX1 - X06SA at the Swiss Light Source synchrotron at 1.0 Å wavelength using DA+ data acquisition and analysis software<sup>34</sup>. Data were integrated and scaled with the XDS program<sup>35</sup>. The search model was generated from the previously published structure of the yeast 80S ribosome<sup>29</sup> (PDB ID 4V88). The initial molecular replacement solution was refined in PHENIX by rigid-body refinement with the 40S and 60S subunits treated as rigid bodies. After initial refinement, density corresponding to the mRNA, tRNAs, eEF2 as well as ribosome rearrangements became clearly visible in the difference electron density map. The crystal structure of eEF2 (PDB ID 1N0U) was docked into the density and manually adjusted before refinement. Refinement was carried out in alternating cycles of automated refinements using PHENIX with manual refinement and model building in COOT resulting in a model with Ramachandran favoured, allowed and outliers of 83.4%, 14.4% and 1.1% respectively. A summary of refinement and data collection statistics is displayed in Extended Data Table 1. All figures were generated using PyMOL.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Coordinates and structure factors have been deposited with the Protein Data Bank under accession code [7OSM](#). Previously published models that were used for analysis and comparison are also available in the Protein Data Bank with accession codes [4V88](#), [4V6F](#), [1N0U](#), [3J77](#), [3J78](#), [4V6F](#), [6GZ3](#) and [6GZ5](#).

## References

1. 1.

Ortiz, P. A., Ulloque, R., Kihara, G. K., Zheng, H. & Kinzy, T. G. Translation elongation factor 2 anticodon mimicry domain mutants affect fidelity and diphtheria toxin resistance. *J. Biol. Chem.* **281**, 32639–32648 (2006).

2. 2.

Waas, W. F., Druzina, Z., Hanan, M. & Schimmel, P. Role of a tRNA base modification and its precursors in frameshifting in eukaryotes. *J. Biol. Chem.* **282**, 26026–26034 (2007).

3. 3.

Noller, H. F., Lancaster, L., Mohan, S. & Zhou, J. Ribosome structural dynamics in translocation: yet another functional role for ribosomal RNA. *Q. Rev. Biophys.* **50**, e12 (2017).

4. 4.

Murray, J. et al. Structural characterization of ribosome recruitment and translocation by type IV IRES. *Elife* **5**, e13567 (2016).

5. 5.

Jørgensen, R., Merrill, A. R. & Andersen, G. R. The life and death of translation elongation factor 2. *Biochem. Soc. Trans.* **34**, 1–6 (2006).

6. 6.

Flis, J. et al. tRNA translocation by the eukaryotic 80S ribosome and the impact of GTP hydrolysis. *Cell Rep.* **25**, 2676–2688.e7 (2018).

7. 7.

Pellegrino, S. et al. Structural insights into the role of diphthamide on elongation factor 2 in mRNA reading-frame maintenance. *J. Mol. Biol.* **430**, 2677–2687 (2018).

8. 8.

Spahn, C. M. et al. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* **23**, 1008–1019 (2004).

9. 9.

Taylor, D. J. et al. Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J.* **26**, 2421–2431 (2007).

10. 10.

Voorhees, R. M., Fernández, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the mammalian ribosome–Sec61 complex to 3.4 Å resolution. *Cell* **157**, 1632–1643 (2014).

11. 11.

Bhatt, P. R. et al. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).

12. 12.

Gavrilova, L. P., Koteliansky, V. E. & Spirin, A. S. Ribosomal protein S12 and ‘non-enzymatic’ translocation. *FEBS Lett.* **45**, 324–328 (1974).

13. 13.

Ogle, J. M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).

14. 14.

Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256–259 (2012).

15. 15.

Liu, S. et al. Diphthamide modification on eukaryotic elongation factor 2 is needed to assure fidelity of mRNA translation and mouse development. *Proc. Natl Acad. Sci. USA* **109**, 13817–13822 (2012).

16. 16.

Ulyanov, N. B. & James, T. L. RNA structural motifs that entail hydrogen bonds involving sugar-phosphate backbone atoms of RNA. *New J. Chem.* **34**, 910–917 (2010).

17. 17.

Khade, P. K. & Joseph, S. Messenger RNA interactions in the decoding center control the rate of translocation. *Nat. Struct. Mol. Biol.* **18**, 1300–1302 (2011).

18. 18.

Prokhorova, I. et al. Aminoglycoside interactions and impacts on the eukaryotic ribosome. *Proc. Natl Acad. Sci. USA* **114**, E10899–E10908 (2017).

19. 19.

Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D. & Steitz, T. A. The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat. Struct. Mol. Biol.* **17**, 289–293 (2010).

20. 20.

Peske, F., Savelsbergh, A., Katunin, V. I., Rodnina, M. V. & Wintermeyer, W. Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J. Mol. Biol.* **343**, 1183–1194 (2004).

21. 21.

Rosselló-Tortella, M. et al. Epigenetic loss of the transfer RNA-modifying enzyme TYW2 induces ribosome frameshifts in colon cancer. *Proc. Natl Acad. Sci. USA* **117**, 20785–20793 (2020).

22. 22.

Carlson, B. A. et al. Transfer RNA modification status influences retroviral ribosomal frameshifting. *Virology* **255**, 2–8 (1999).

23. 23.

Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).

24. 24.

Cornish, P. V., Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell* **30**, 578–588 (2008).

25. 25.

Salsi, E., Farah, E. & Ermolenko, D. N. EF-G activation by phosphate analogs. *J. Mol. Biol.* **428**, 2248–2258 (2016).

26. 26.

Susorov, D. et al. Eukaryotic translation elongation factor 2 (eEF2) catalyzes reverse translocation of the eukaryotic ribosome. *J. Biol. Chem.* **293**, 5220–5229 (2018).

27. 27.

Chen, C. et al. Elongation factor G initiates translocation through a power stroke. *Proc. Natl Acad. Sci. USA* **113**, 7515–7520 (2016).

28. 28.

Belardinelli, R. et al. Choreography of molecular movements during ribosome progression along mRNA. *Nat. Struct. Mol. Biol.* **23**, 342–348 (2016).

29. 29.

Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209 (2010).

30. 30.

Jørgensen, R., Carr-Schmid, A., Ortiz, P. A., Kinzy, T. G. & Andersen, G. R. Purification and crystallization of the yeast elongation factor eEF2. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 712–715 (2002).

31. 31.

Spirin, A. S., Belitsina, N. V. & Yusupova, G. Z. Ribosomal synthesis of polypeptides from aminoacyl-tRNA without polynucleotide template. *Methods Enzymol.* **164**, 631–649 (1988).

32. 32.

Mesters, J. R., Vorstenbosch, E. L. H., de Boer, A. J. & Kraal, B. Complete purification of tRNA, charged or modified with hydrophobic groups, by reversed-phase high-performance liquid chromatography on a C4/C18 column system. *J. Chromatogr. A* **679**, 93–98 (1994).

33. 33.

Mechulam, Y., Guillon, L., Yatime, L., Blanquet, S. & Schmitt, E. Protection-based assays to measure aminoacyl-tRNA binding to translation initiation factors. *Methods Enzymol.* **430**, 265–281 (2007).

34. 34.

Wojdyla, J. A. et al. DA+ data acquisition and analysis software at the Swiss Light Source macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **25**, 293–303 (2018).

35. 35.

Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).

36. 36.

Afonine, P. V. et al. FEM: feature-enhanced map. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 646–666 (2015).

37. 37.

Zhou, J., Lancaster, L., Donohue, J. P. & Noller, H. F. Spontaneous ribosomal translocation of mRNA and tRNAs into a chimeric hybrid state. *Proc. Natl Acad. Sci. USA* **116**, 7813–7818 (2019).

## Acknowledgements

We thank the staff of the PXI beamline at the Swiss Light Source (Switzerland) and M. Wang for assistance during synchrotron data collection; N. Milicevic for the preparation of aminoacylated tRNAs samples; and I. Prokhorova for assistance with purification of the 80S ribosome and eEF2. Grant ANR-10-LABX-0030-INRT from a French State Fund managed by the ANR under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02 (to M.D.). M.D. started working on the project as an IGBMC International PhD Programme fellow supported by LabEx INRT funds. This

work was supported by the French National Research Agency ANR-15-CE11-0021-01 (to G.Y.) and La Fondation pour la Recherche Médicale DEQ20181039600 (to M.Y.).

## Author information

### Author notes

1. Natalia Demeshkina

Present address: Biochemistry and Biophysics Center, National Heart, Lung and Blood Institute, Bethesda, MD, USA

### Affiliations

1. Institute of Genetics and Molecular and Cellular Biology, CNRS UMR7104, INSERM U1258, University of Strasbourg, Illkirch, Strasbourg, France

Muminjon Djumagulov, Natalia Demeshkina, Lasse Jenner, Alexey Rozov, Marat Yusupov & Gulnara Yusupova

2. Urania Therapeutics, Ostwald, France

Muminjon Djumagulov & Alexey Rozov

### Contributions

M.D. and N.D. conducted biochemical and crystallization experiments. M.D., N.D., L.J. and A.R. participated in data collection. M.D., L.J. and A.R. performed molecular model building and refinement. Main data analysis was performed by M.D. N.D., L.J. and A.R. also contributed to the data analysis. M.Y. and G.Y. conceived and supervised the project. All authors discussed the results and commented on the manuscript.

### Corresponding authors

Correspondence to [Marat Yusupov](#) or [Gulnara Yusupova](#).

### Ethics declarations

### Competing interests

The authors declare no competing interests

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Examples of the electron density maps of the 80S ribosome translocation intermediate complex.](#)

FEM<sup>36</sup> map for a representative part of 25S rRNA (**a**) and for eEF2 (**b**) contoured at 2 sigma.  $2F_o - F_c$  electron density map contoured at 1.5 sigma for 25S rRNA (**c**) and at 1 sigma for the diphthamide loop of eEF2 (**d**).

### [Extended Data Fig. 2 Comparison of various states of tRNAs on the ribosome.](#)

**a**, Positions of tRNAs of the intermediate translocation complex relative to tRNAs in the classical A, P and E states (grey, PDB ID 4V6F) aligned on the SSU body; indicated distances are in Å. **b**, Positions of tRNAs relative to tRNAs in the classical nonrotated state (PDB ID 3J78) aligned on LSU. **c**, Chemical structure of the wybutosine tRNA<sup>Phe</sup> modification at position 37.

### [Extended Data Fig. 3 Insertion of eEF2 into the SSU A site in the late translocation complex TI-POST1 \(yellow, PDB ID 6GZ3\) in comparison to the position of eEF2 in the early translocation intermediate state that we determined in this study \(red\).](#)

View from the LSU side. Alignment is performed using the SSU body.

### [Extended Data Fig. 4 Interactions of eEF2 with the SSU body.](#)

**a**, Comparison of the current X-ray structure of an early translocation intermediate (18S rRNA in cyan) with cryo-EM models of late translocation steps TI-POST1 (magenta, PDB ID 6GZ3) and TI-POST3 (grey, PDB ID 6GZ3)<sup>6</sup>. Superimposition of

eEF2 in the two structures shows that there is a noticeable shift of 18S rRNA of the SSU body away from eEF2 at later stages of translocation. **b**, Contacts between eEF2 and h5 of SSU of the rotated ribosome in the reported intermediate translocation complex. **c**, Interactions between eEF2 and the SSU shoulder proteins eS30 and uS12. Stabilization of the eEF2 domain IV by the N-terminus of eukaryote-specific protein eS30 that itself interacts with conserved decoding protein uS12. Presumably, eS30 co-evolved with eEF2, whose domain IV has 65 additional amino acids compared to its bacterial counterpart EF-G<sup>5</sup>, to provide supplementary stabilization as well as to enhance propagation of conformational changes at the decoding site<sup>13,14</sup>. **d**, A close up view of the dashed region from (c). Movement of uS12 (arrow 1) induced by the SSU body back-rotation can propagate to switch II (in cyan) of eEF2 through domain III (arrow 2) and can trigger GTP hydrolysis or Pi release. Regions of uS12 adjacent to eEF2 are shown and coloured dark- to light-green based on the rotational state where dark-green is the most rotated state (intermediate translocation complex) and medium (TI-POST1) and light (TI-POST3) greens represent the least rotated states. Alignment was done using eEF2.

**Extended Data Fig. 5 Rearrangement of h34 and h31 of SSU with ASL movement of tRNAs in the translocation intermediate complex.**

**a**, A new position of conserved C1274 in h34 of 18S rRNA (C1054 of 16S rRNA in bacteria, Fig. 1d) relative to ASL of the A site tRNA and mRNA. By establishing a contact with mRNA nucleotide in position (+7) C1274 can contribute to maintaining mRNA reading frame during translocation. **b**, In the early intermediate state of translocation h31 of 18S rRNA moves together with pe/E tRNA. In this state, the U1191-C1637 interaction is broken, however, U1191 remains in contact with C35 of pe/E tRNA. A similar situation was described for the bacterial ribosome with the pe/E tRNA in a structure modelling spontaneous translocation without EF-G<sup>37</sup>. Compared to the intermediate eukaryotic state reported here, the bacterial ap/P tRNA coupled to mRNA is shifted towards the P-site indicating translational reading frame slippage in the factor-free system. This comparison shows the importance of role that translocase eEF2 (or EF-G) fulfils in the coordinated movement of tRNA-mRNA during translocation.

**Extended Data Fig. 6 Contacts between eEF2, tRNAs and the 80S ribosome in the crystal structure of the translocation intermediate complex.**

**a**, Position of diphthamide at the conjunction of ap/P tRNA<sup>Phe</sup>, mRNA and decoding adenines of h44 and interactions of eEF2 (Ala652) with rRNA of the SSU head (U1442 of 18S). **b**, A solvent view on the P stalk region. **c**, Movement of the P stalk

upon eEF2 binding seen when comparing our structure to that of the vacant 80S ribosome structure (PDB ID 4V88). Alignment was performed using 25S rRNA as a reference. **d**, The G domain of eEF2 with ordered switch I and II regions indicating that this is a pre-hydrolysis state (left panel). Close-up view of the GTP pocket and sarcin-ricin loop (SRL) of 25S rRNA (right panel). **e**, Disruption of the B1a bridge consisting of helix 38 of 25S rRNA (A-site finger) and protein uS13 induced by rotation of the SSU head and body. Non-rotated 80S (PDB ID 3J78) is coloured in black. **f**, Stacking interactions of rRNA elements (magnified in the right panel) of the L1 stalk with the elbow region of pe/E tRNA.

### [Extended Data Fig. 7 Comparison of secondary structures of the decoding loop in bacteria \(left\) and eukaryotes \(right\).](#)

In bacteria the internal loop of h44 of 16S rRNA consists of two nucleotides (A1492–A1493) on the 3'-side and has one nucleotide on the 5'-side (A1408). In contrast, eukaryotic 18S rRNA contains at least one additional nucleotide on each side (dashed box). Secondary structure diagrams of helix 44 from bacteria (16S rRNA, left) and from yeast and human (18S rRNA, right).

### **Extended Data Table 1 Data collection and refinement statistics**

## **Supplementary information**

### [Reporting Summary](#)

### [Supplementary Video 1](#)

Translocation of tRNAs on the 40S ribosome subunit and codon–anticodon stabilization by diphthamide and wybutosine. Intersubunit movement in the 80S ribosome and internal conformational changes of the 40S ribosomal subunit during eEF2 catalysed translocation of tRNAs. The movie is based on the presented crystal structure of early translocation-intermediate complex (PDB ID 7OSM), as well as cryo-EM structures of late translocation (PDB ID 6GZ3, 6GZ5).

## **Rights and permissions**

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## [Reprints and Permissions](#)

## About this article

### Cite this article

Djumagulov, M., Demeshkina, N., Jenner, L. *et al.* Accuracy mechanism of eukaryotic ribosome translocation. *Nature* **600**, 543–546 (2021). <https://doi.org/10.1038/s41586-021-04131-9>

- Received: 24 June 2021
- Accepted: 13 October 2021
- Published: 01 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04131-9>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04131-9>

- Article
- [Published: 01 December 2021](#)

# De novo protein design by deep network hallucination

- [Ivan Anishchenko](#) ORCID: [orcid.org/0000-0003-3645-2044](#)<sup>1,2</sup> nat,
- [Samuel J. Pellock](#) ORCID: [orcid.org/0000-0002-7557-7985](#)<sup>1,2</sup> nat,
- [Tamuka M. Chidyausiku](#) ORCID: [orcid.org/0000-0002-2561-782X](#)<sup>1,2</sup>,
- [Theresa A. Ramelot](#) ORCID: [orcid.org/0000-0002-0335-1573](#)<sup>3,4</sup>,
- [Sergey Ovchinnikov](#) ORCID: [orcid.org/0000-0003-2774-2744](#)<sup>5</sup>,
- [Jingzhou Hao](#)<sup>3,4</sup>,
- [Khushboo Bafna](#) ORCID: [orcid.org/0000-0002-1893-1175](#)<sup>3,4</sup>,
- [Christoffer Norn](#) ORCID: [orcid.org/0000-0002-1450-4651](#)<sup>1,2</sup>,
- [Alex Kang](#)<sup>1,2</sup>,
- [Asim K. Bera](#) ORCID: [orcid.org/0000-0001-9473-2912](#)<sup>1,2</sup>,
- [Frank DiMaio](#)<sup>1,2</sup>,
- [Lauren Carter](#)<sup>1,2</sup>,
- [Cameron M. Chow](#) ORCID: [orcid.org/0000-0001-5351-6412](#)<sup>1,2</sup>,
- [Gaetano T. Montelione](#) ORCID: [orcid.org/0000-0002-9440-3059](#)<sup>3,4</sup> &
- [David Baker](#) ORCID: [orcid.org/0000-0001-7896-6217](#)<sup>1,2,6</sup>

[Nature](#) volume 600, pages 547–552 (2021)

- 14k Accesses
- 1 Citations
- 389 Altmetric
- [Metrics details](#)

## Subjects

- [Machine learning](#)
- [Protein design](#)
- [Protein folding](#)

## Abstract

There has been considerable recent progress in protein structure prediction using deep neural networks to predict inter-residue distances from amino acid sequences<sup>1,2,3</sup>. Here we investigate whether the information captured by such networks is sufficiently rich to generate new folded proteins with sequences unrelated to those of the naturally occurring proteins used in training the models. We generate random amino acid sequences, and input them into the trRosetta structure prediction network to predict starting residue–residue distance maps, which, as expected, are quite featureless. We then carry out Monte Carlo sampling in amino acid sequence space, optimizing the contrast (Kullback–Leibler divergence) between the inter-residue distance distributions predicted by the network and background distributions averaged over all proteins. Optimization from different random starting points resulted in novel proteins spanning a wide range of sequences and predicted structures. We obtained synthetic genes encoding 129 of the network-‘hallucinated’ sequences, and expressed and purified the proteins in *Escherichia coli*; 27 of the proteins yielded monodisperse species with circular dichroism spectra consistent with the hallucinated structures. We determined the three-dimensional structures of three of the hallucinated proteins, two by X-ray crystallography and one by NMR, and these closely matched the hallucinated models. Thus, deep networks trained to predict native protein structures from their sequences can be inverted to design new proteins, and such networks and methods should contribute alongside traditional physics-based models to the de novo design of proteins with new functions.

This is a preview of subscription content

## Access options

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

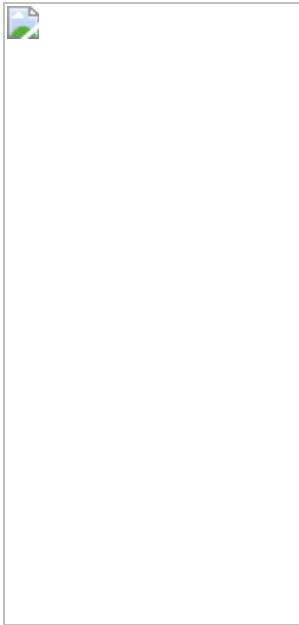
[Rent or Buy](#)

All prices are NET prices.

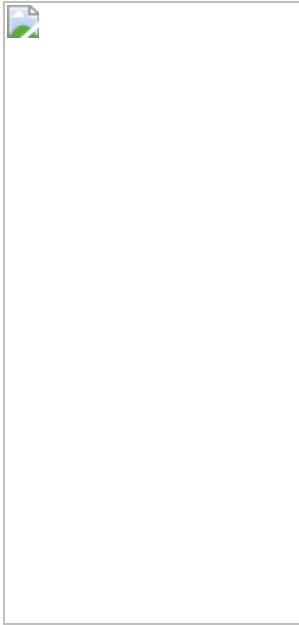
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

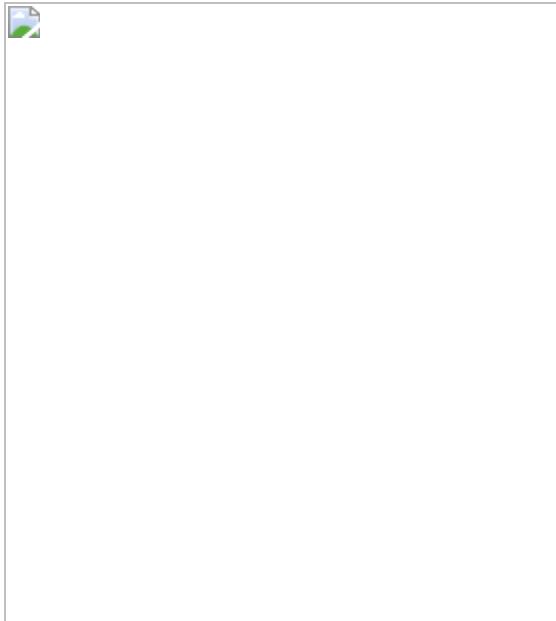
**Fig. 1: Overview of protein hallucination approach.**



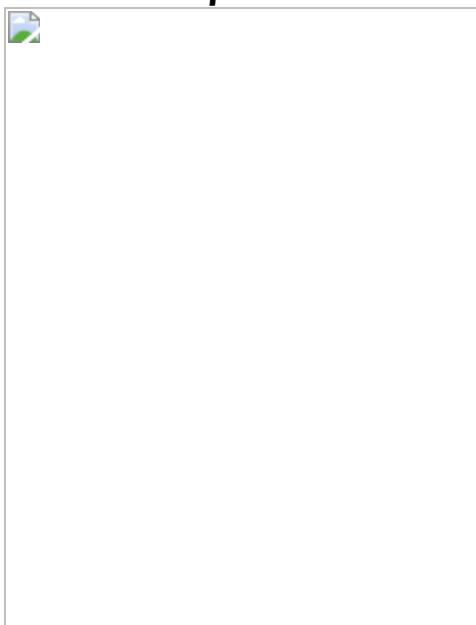
**Fig. 2: Overview of computational results.**



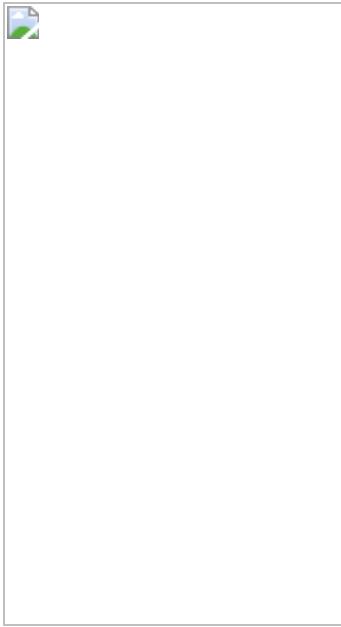
**Fig. 3: Experimental characterization of  $\alpha$ -helical network-hallucinated proteins.**



**Fig. 4: Experimental characterization of network-hallucinated proteins with mixed  $\alpha$ - $\beta$  structures.**



**Fig. 5: Structural analysis of network-hallucinated proteins.**



## Data availability

The atomic coordinates of the crystal structures for designs 0217 and 0738\_mod, as well as the NMR structure for design 0515 have been deposited in the RCSB Protein Data Bank with the accession numbers [7K3H](#), [7M0Q](#) and [7M5T](#), respectively. NMR chemical shifts, NOESY peak lists, and spectral data have been deposited in the BioMagResDB, BMRB ID 30890. Amino acid sequences and structure models for all 2K designs described in the manuscript are freely available for download at <https://files.ipd.uw.edu/pub/trRosetta/hallucinations2K.tar.gz>. Amino acid sequences and 3D structures of the generated designs were compared to known protein sequences and structures in UniProt ([https://ftp.uniprot.org/pub/databases/uniprot/previous\\_releases/release-2017\\_12/uniref/](https://ftp.uniprot.org/pub/databases/uniprot/previous_releases/release-2017_12/uniref/)) and the Protein Data Bank (11 March 2020), respectively.

## Code availability

The computer code used to generate the hallucinated proteins described in the manuscript was made publicly available as a part of trDesign Github package (<https://github.com/gjoni/trDesign>); corresponding structural models were generated by the trRosetta structure modelling script available for free download at <https://yanglab.nankai.edu.cn/trRosetta/download/>.

The Rosetta software suite was used to perform ab initio prediction calculations. Rosetta is freely available for academic users on Github, and can be licensed for commercial use by the University of Washington CoMotion Express License Program.

## References

1. 1.  
Xu, J. Distance-based protein folding powered by deep learning. *Proc. Natl Acad. Sci. USA* **116**, 16856–16865 (2019).
2. 2.  
Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
3. 3.  
Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
4. 4.  
Biswas, S., Khimulya, G., Alley, E. C., Esvelt, K. M. & Church, G. M. Low- $N$  protein engineering with data-efficient deep learning. *Nat. Methods* **18**, 389–396 (2021).
5. 5.  
Madani, A. et al. ProGen: language modeling for protein generation. Preprint at <https://arxiv.org/abs/2004.03497> (2020).
6. 6.  
Anand, N., Eguchi, R. & Huang, P. S. Fully differentiable full-atom protein backbone generation. In *ICLR 2019 Workshop*

<https://openreview.net/forum?id=SJxnVL8YOV> (2019).

7. 7.

Wang, J., Cao, H., Zhang, J. Z. H. & Qi, Y. Computational protein design with deep learning neural networks. *Sci Rep.* **8**, 6349 (2018).

8. 8.

Ingraham, J., Garg, V. K., Barzilay, R. & Jaakkola, T. Generative models for graph-based protein design. in *ICLR 2019 Workshop* <https://openreview.net/forum?id=SJgxrlLKOE> (2019).

9. 9.

Anand, N., Eguchi, R. R., Derry, A., Altman, R. B. & Huang, P.-S. Protein sequence design with a learned potential. Preprint at <https://doi.org/10.1101/2020.01.06.895466> (2020).

10. 10.

Strokach, A., Becerra, D., Corbi-Verge, C., Perez-Riba, A. & Kim, P. M. Fast and flexible protein design using deep graph neural networks. *Cell Syst.* **11**, 402–411.e4 (2020).

11. 11.

Karimi, M., Zhu, S., Cao, Y. & Shen, Y. De novo protein design for novel folds using guided conditional Wasserstein generative adversarial networks. *J. Chem. Inf. Model.* **60**, 5667–5681 (2020).

12. 12.

Davidsen, K. et al. Deep generative models for T cell receptor protein sequences. *eLife* **8**, e46935 (2019).

13. 13.

Costello, Z. & Martin, H. G. How to hallucinate functional proteins. Preprint at <https://arxiv.org/abs/1903.00458> (2019).

14. 14.

Eguchi, R. R., Anand, N., Choe, C. A. & Huang, P.-S. IG-VAE: generative modeling of immunoglobulin proteins by direct 3D coordinate generation. Preprint at <https://doi.org/10.1101/2020.08.07.242347> (2020).

15. 15.

Repecka, D. et al. Expanding functional protein sequence spaces using generative adversarial networks. *Nat. Mach. Intell.* **3**, 324–333 (2021).

16. 16.

Hawkins-Hooker, A. et al. Generating functional protein variants with variational autoencoders. *PLoS Comput. Biol.* **17**, e1008736 (2021).

17. 17.

Senior, A. W. et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins* **87**, 1141–1148 (2019).

18. 18.

Mordvintsev, A., Olah, C. & Tyka, M. Inceptionism: going deeper into neural networks. *Google AI Blog* <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html> (2015).

19. 19.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

20. 20.

Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).

21. 21.

Park, H. et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

22. 22.

Rossi, P. et al. A microscale protein NMR sample screening pipeline. *J. Biomol. NMR* **46**, 11–22 (2010).

23. 23.

Koga, N. et al. Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).

24. 24.

Dou, J. et al. De novo design of a fluorescence-activating  $\beta$ -barrel. *Nature* **561**, 485–491 (2018).

25. 25.

Norn, C. et al. Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. USA* **118**, e2017228118 (2021).

26. 26.

Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

27. 27.

Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

28. 28.

Wang, J. et al. Deep learning methods for designing proteins scaffolding functional sites. Preprint at <https://doi.org/10.1101/2021.11.10.468128> (2021).

29. 29.

Jendrusch, M., Korbel, J. O. & Sadiq, S. K. AlphaDesign: A de novo protein design framework based on AlphaFold. Preprint at <https://doi.org/10.1101/2021.10.11.463937> (2021).

30. 30.

Tischer, D. et al. Design of proteins presenting discontinuous functional sites using deep learning. Preprint at <https://doi.org/10.1101/2020.11.29.402743> (2020).

31. 31.

Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

32. 32.

Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

33. 33.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423 (1995).

34. 34.

Acton, T. B. et al. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* **493**, 21–60 (2011).

35. 35.

Xiao, R. et al. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol.* **172**, 21–33 (2010).

36. 36.

Jansson, M. et al. High-level production of uniformly <sup>15</sup>N-and <sup>13</sup>C-enriched fusion proteins in *Escherichia coli*. *J. Biomol. NMR* **7**, 131–141 (1996).

37. 37.

Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).

38. 38.

Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

39. 39.

Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

40. 40.

Favier, A. & Brutscher, B. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR* **73**, 199–211 (2019).

41. 41.

Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. *J. Biomol. NMR* **52**, 315–327 (2012).

42. 42.

Ying, J., Delaglio, F., Torchia, D. A. & Bax, A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **68**, 101–118 (2017).

43. 43.

Lee, W. et al. I-PINE web server: an integrative probabilistic NMR assignment system for proteins. *J. Biomol. NMR* **73**, 213–222 (2019).

44. 44.

Moseley, H. N. B., Sahota, G. & Montelione, G. T. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. *J. Biomol. NMR* **28**, 341–355 (2004).

45. 45.

Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).

46. 46.

Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).

47. 47.

Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002).

48. 48.

Huang, Y. J., Powers, R. & Montelione, G. T. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.* **127**, 1665–1674 (2005).

49. 49.

Huang, Y. J., Tejero, R., Powers, R. & Montelione, G. T. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins* **62**, 587–603 (2006).

50. 50.

Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

51. 51.

Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778–795 (2007).

52. 52.

Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

53. 53.

McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

54. 54.

DiMaio, F. et al. Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10**, 1102–1104 (2013).

55. 55.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

56. 56.

Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

57. 57.

Theobald, D. L. & Wuttke, D. S. Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput. Biol.* **4**, e43 (2008).

58. 58.

*The PyMOL Molecular Graphics System* version 2.4 (Schrödinger, 2021).

59. 59.

Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–690 (2008).

60. 60.

Montelione, G. T. & Wagner, G. 2D Chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *J. Am. Chem. Soc.* **111**, 3096–3098 (1989).

## Acknowledgements

We thank R. Xiao, G. Liu and A. Wu (Nexomics Biosciences), for assistance in initial NMR protein production; J. Aramini for assistance with NMR data collection for initial HSQC screening; R. Ballard and X. Li for mass spectrometry assistance; and R. Divine and R. Kibler for AKTA scripting. This work was funded by grants from the NSF (DBI 1937533 to D.B. and I.A., and MCB 2032259 to S.O.), the NIH (DP5OD026389 to S.O.), Open Philanthropy (C.C. and A.B.), Eric and Wendy Schmidt by recommendation of the Schmidt Futures program (F.D. and L.C.), and the Audacious project (A.K.), the Washington Research Foundation (S.J.P.), Novo Nordisk Foundation Grant NNF17OC0030446 (C.N.). This work was also supported in part by NIH grants R01 GM120574 (G.T.M.) and R35GM141818 (G.T.M.), and the Howard Hughes Medical Institute (D.B. and T.M.C.). We also acknowledge computing resources provided by the Hyak supercomputer system funded by the STF at the University of Washington, and Rosetta@Home volunteers in ab initio structure prediction calculations, and thank staff at Northeastern Collaborative Access Team at Advanced Photon Source for the beamline, supported by NIH grants P30GM124165 and S10OD021527, and DOE contract DE-AC02-06CH11357. We acknowledge the NMR Core Facility resources at Rensselaer Polytechnic Institute and thank S. McCallum for providing valuable support.

## Author information

### Author notes

1. These authors contributed equally: Ivan Anishchenko, Samuel J. Pellock

## Affiliations

1. Department of Biochemistry, University of Washington, Seattle, WA, USA

Ivan Anishchenko, Samuel J. Pellock, Tamuka M.  
Chidyausiku, Christoffer Norn, Alex Kang, Asim K. Bera, Frank  
DiMaio, Lauren Carter, Cameron M. Chow & David Baker

2. Institute for Protein Design, University of Washington, Seattle, WA,  
USA

Ivan Anishchenko, Samuel J. Pellock, Tamuka M.  
Chidyausiku, Christoffer Norn, Alex Kang, Asim K. Bera, Frank  
DiMaio, Lauren Carter, Cameron M. Chow & David Baker

3. Department of Chemistry and Chemical Biology, Rensselaer  
Polytechnic Institute, Troy, NY, USA

Theresa A. Ramelot, Jingzhou Hao, Khushboo Bafna & Gaetano T.  
Montelione

4. Center for Biotechnology and Interdisciplinary Sciences, Rensselaer  
Polytechnic Institute, Troy, NY, USA

Theresa A. Ramelot, Jingzhou Hao, Khushboo Bafna & Gaetano T.  
Montelione

5. John Harvard Distinguished Science Fellowship Program, Harvard  
University, Cambridge, MA, USA

Sergey Ovchinnikov

6. Howard Hughes Medical Institute, University of Washington, Seattle,  
WA, USA

David Baker

## Corresponding author

Correspondence to [David Baker](#).

## Ethics declarations

## Competing interests

G.T.M. is a founder of Nexomics Biosciences. The other authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Comparison of the hallucinated designs to proteins with known structure and of similar length \(100 +/- 10 aa\) from the trRosetta training set.](#)

**a,b)** Multidimensional scaling plots of the sequence (**a**) and structure (**b**) spaces covered by the 2,000 hallucinated proteins (blue dots) along with 1,110 proteins of similar length from the trRosetta training set (red dots). These scatter plots show that subspaces spanned by hallucinated proteins and natural proteins of similar size (100  $\pm$  10 aa) are quite distinct; the network is not simply recapitulating native proteins of the same length. Soluble and structurally characterized hallucinations are marked by black and magenta dots respectively. **c,d)** Distributions of pairwise structure (**c**) and sequence (**d**) similarities for hallucinated and natural proteins. The hallucinated proteins are more similar to each other (blue lines) than they are to natural proteins (grey lines). **e)** Sequence comparisons (gappless threading) of fragments of various size (15,20,...,60 aa) from the hallucinated designs (blue) and natural 100 ( $\pm$  10) aa-long proteins (red) to other proteins from the trRosetta training set. There is no apparent tendency for the trRosetta-based design procedure to “copy over” sequence fragments from the proteins in the training set into the hallucinated designs.

**f,g)** Secondary structure content of the hallucinated designs and natural 100 aa-long proteins from the training set. Hallucinations are more ideal than natural proteins in having less loops but longer secondary structure elements.

**Extended Data Fig. 2 Additional data on the experimentally characterized all- $\alpha$  and mixed  $\alpha$ - $\beta$  network-hallucinated proteins.**

**a,e)** Dendrograms showing representative hallucinated protein designs clustered by TM-score; thermostable designs with CD spectra consistent with the target structure are labelled by their IDs. **b,f)** Three-dimensional models of the hallucinated designs. **c,g)** Predicted distance maps at the end of the hallucination trajectory. **d,h)** Temperature dependence of CD signal at 220 nm in the 25-95 °C temperature range.

**Extended Data Fig. 3 Additional examples of thermostable hallucinations with CD spectra consistent with the target structure.**

**a,g)** 3D structure models of the hallucinated designs. **b,h)** Predicted distance maps at the end of the hallucination trajectory. **c,i)** ab initio folding funnels from Rosetta. **d,j)** Size-exclusion chromatography traces. **e,k)** Circular dichroism spectra at 25 °C (blue) and 95 °C (red). **f,l)** Temperature dependence of Circular Dichroism signal at 220 nm in the 25 to 95 °C temperature range.

**Extended Data Fig. 4 Comparison of 0515 NMR structure to hallucinated model.**

**a)** Superposition of hallucinated model (blue) and NMR medoid structure (gray) of 0515 reveal 1.82 Å backbone r.m.s.d. over 100 residues **b)** Hallucinated model of 0515 colored by distance between C $\alpha$ -C $\alpha$  pairs between model and NMR medoid structure after structural superposition and **b)** corresponding plot of per-residue C $\alpha$ -C $\alpha$  distance difference between model and NMR medoid structure.

## Extended Data Fig.5 Structural analysis of 0217 and comparison to hallucinated model.

**a)** Representative electron density ( $2\text{Fo}-\text{Fc}$ ,  $1\sigma$ ) over entire asymmetric unit (left) and core packing regions (right) of hallucination 0217. **b)** Both chains of the crystal structure colored by B-factor. **c)** Structural superposition of chains observed in the asymmetric unit reveal a  $2.8 \text{ \AA}$  backbone r.m.s.d. over 91 residues. **d)** Crystal lattice contacts for chain A (green) and chain B (yellow) may explain structural differences observed between chains. Circled regions highlight where chain A is an ordered helix-loop-helix and chain B is disordered. **e)** Hallucinated model of 0217 colored by distance between  $\text{Ca}-\text{Ca}$  pairs between model and crystal structure after structural superposition and corresponding plot of per-residue  $\text{Ca}-\text{Ca}$  distance difference between model and crystal structure. **f)** Structural superposition of the hallucinated model and chain B of the 0217 crystal structure (left), 0217 model colored by  $\text{Ca}-\text{Ca}$  distance between hallucination and crystal structure (middle), and per residue  $\text{Ca}-\text{Ca}$  distance between hallucination and crystal structure per residue (right).

## Extended Data Fig. 6 Structural analysis, NMR characterization, and SEC analysis of hallucinated sequence 0417.

**a)** Hallucinated model with surface hydrophobics shown as sticks and **b)** [ $^1\text{H}-^{15}\text{N}$ ]-SOFAST-HMQC spectra of hallucinated sequence 0417 before (red) and after (blue) buffer optimization. Spectrum before optimization (red) was obtained using a protein concentration of  $\sim 0.3 \text{ mM}$  at  $298\text{K}$  in  $20 \text{ mM}$  Tris-HCl, pH 7.2,  $100 \text{ mM}$  NaCl and spectrum acquired after optimization (blue) was obtained using a protein concentration of  $\sim 0.3 \text{ mM}$ , at temperature of  $323 \text{ K}$  in a buffer of  $20 \text{ mM}$  sodium phosphate at pH 6.5,  $50 \text{ mM}$  NaCl, and 20% glycerol. The NMR data are consistent with a folded structure containing a mix of alpha and beta secondary structure. Even under optimized conditions, there is still evidence of exchange broadening (e.g. Trp side chain  $\text{N}^\epsilon\text{Hs}$  are weak), resonances that appear only at high temperature and high glycerol concentrations, and some resonances that are doubled; all indications of transient self-association. **c)**

Size-exclusion chromatography trace of 0417 displays a small additional peak corresponding to a larger oligomeric species which corroborates the NMR analysis.

**Extended Data Fig. 7 Structural analysis of 0738\_mod and comparison to hallucinated model 0738.**

**a**) Representative electron density ( $2\text{Fo}-\text{Fc}$ ,  $1\sigma$ ) over entire asymmetric unit (left) and core packing regions (right) of hallucination 0738\_mod. **b**) Both chains of the crystal structure colored by B-factor. **c**) Structural superposition of the hallucinated model and chain A of the 0738\_mod crystal structure (left), 0738\_mod model colored by  $\text{Ca}-\text{Ca}$  distance between hallucination and crystal structure (middle), and per residue  $\text{Ca}-\text{Ca}$  distance between hallucination and crystal structure per residue (right). **d**) Hallucinated model of 0738\_mod colored by distance between  $\text{Ca}-\text{Ca}$  pairs between model and crystal structure after structural superposition and corresponding plot of per-residue  $\text{Ca}-\text{Ca}$  distance difference between model and crystal structure.

**Extended Data Fig. 8 NMR and biochemical analysis of hallucinated sequences 0515, 0738\_mod, and 0217.**

**a**)  $^1\text{H}-^{15}\text{N}$  heteronuclear NOE (hetNOE) histograms for 0515 (82 non-overlapped peaks), 0738\_mod (144 peaks), and 0217 (47 peaks), together with their average values.  $^1\text{H}-^{15}\text{N}$  steady state heteronuclear NOEs were obtained from the ratio of cross peak intensities ( $I_{\text{saturated}}/I_{\text{equilibrium}}$ ) with ( $I_{\text{saturated}}$ ) and without ( $I_{\text{equilibrium}}$ ) 3 s of proton saturation during the presat delay and recorded in an interleaved manner, split in TopSpin, processed identically using NMRPipe, and peak picked in SPARKY to obtain peak intensities. **b**)  $^1\text{H}-^{15}\text{N}$  HSQC spectra of corresponding proteins collected at 800 MHz at 298 K in 25 mM HEPES, pH 7.4, 50 mM NaCl buffer and prepared in a 5-mm Shigemi NMR tubes for data collection with addition of 5%  $\text{D}_2\text{O}$  (v/v). These  $^{15}\text{N}$ -enriched protein samples were prepared at concentrations of 0.4 mM, 0.15 mM, and 0.2 mM, respectively. **c**) SEC data demonstrating monodispersity of these proteins in solution, with

predominantly monomer for 0515 and 0738\_mod and predominantly dimer for 0217. SDS-PAGE data (not shown) show that each is > 95% homogeneous, which together with MALDI-TOF mass spectrometry indicate that the spectral heterogeneity observed is not due to chemical heterogeneity. **d)** Ribbon diagrams of the corresponding monomeric or dimeric protein structures. These results show that the three designs have characteristic dynamics in solution. The average hetNOE for the homodimer 0217 is lower than for 0515 and 0738\_mod, and it has fewer peaks than expected due to exchange broadening. Although 0738\_mod has a similar hetNOE distribution as monomeric 0515, it has more than double the expected number of peaks, indicating at least two folded conformations (for all or parts of the protein) in solution that are in slow conformational exchange on the NMR time-scale. This was further validated by the appearance of new peaks in spectra at lower temperature (288 K), and different peaks at higher temperatures (308 and 318 K), and confirmed by detection of <sup>15</sup>N ZZ-exchange cross peaks at 318 K with 600 and 750 ms mixing times (Bruker pulse sequence hsqcetxf3gp, data not shown)<sup>60</sup>.

**Extended Data Table 1 NMR refinement statistics and quality scores for 0515**

**Extended Data Table 2 Crystallographic data collection and refinement statistics**

## Supplementary information

### Supplementary Information

This file contains a Supplementary Discussion, Supplementary Table 1 and Supplementary Figs 1–7.

### Reporting Summary

### Peer Review File

## Rights and permissions

## [Reprints and Permissions](#)

## About this article

### Cite this article

Anishchenko, I., Pellock, S.J., Chidyausiku, T.M. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).  
<https://doi.org/10.1038/s41586-021-04184-w>

- Received: 18 September 2020
- Accepted: 21 October 2021
- Published: 01 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04184-w>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04184-w>

- Article
- [Published: 25 October 2021](#)

# The conformational cycle of prestin underlies outer-hair cell electromotility

- [Navid Bavi<sup>1</sup>](#),
- [Michael David Clark<sup>1</sup>](#),
- [Gustavo F. Contreras<sup>1</sup>](#),
- [Rong Shen<sup>1</sup>](#),
- [Bharat G. Reddy<sup>1 nAff3</sup>](#),
- [Wieslawa Milewski<sup>1</sup>](#) &
- [Eduardo Perozo ORCID: orcid.org/0000-0001-7132-2793<sup>1,2</sup>](#)

[Nature](#) volume **600**, pages 553–558 (2021)

- 3898 Accesses
- 38 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Hair cell](#)

## Abstract

The voltage-dependent motor protein prestin (also known as SLC26A5) is responsible for the electromotive behaviour of outer-hair cells and underlies the cochlear amplifier<sup>1</sup>. Knockout or impairment of prestin causes severe hearing loss<sup>2,3,4,5</sup>. Despite the key role of prestin in hearing, the mechanism by which mammalian prestin senses voltage and transduces it into cellular-scale movements (electromotility) is poorly understood. Here we determined the structure of dolphin prestin in six distinct states using single-particle cryo-electron microscopy. Our structural and functional data suggest that prestin adopts a unique and complex set of states, tunable by the identity of bound anions ( $\text{Cl}^-$  or  $\text{SO}_4^{2-}$ ). Salicylate, a drug that can cause reversible hearing loss, competes for the anion-binding site of prestin, and inhibits its function by immobilizing prestin in a new conformation. Our data suggest that the bound anion together with its coordinating charged residues and helical dipole act as a dynamic voltage sensor. An analysis of all of the anion-dependent conformations reveals how structural rearrangements in the voltage sensor are coupled to conformational transitions at the protein–membrane interface, suggesting a previously undescribed mechanism of area expansion. Visualization of the electromotility cycle of prestin distinguishes the protein from the closely related SLC26 anion transporters, highlighting the basis for evolutionary specialization of the mammalian cochlear amplifier at a high resolution.

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.  
VAT will be added later in the checkout.  
Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

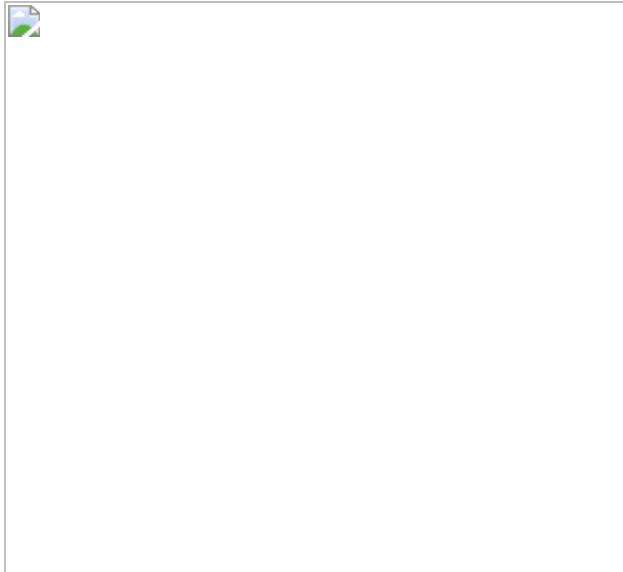
[Rent or Buy](#)

All prices are NET prices.

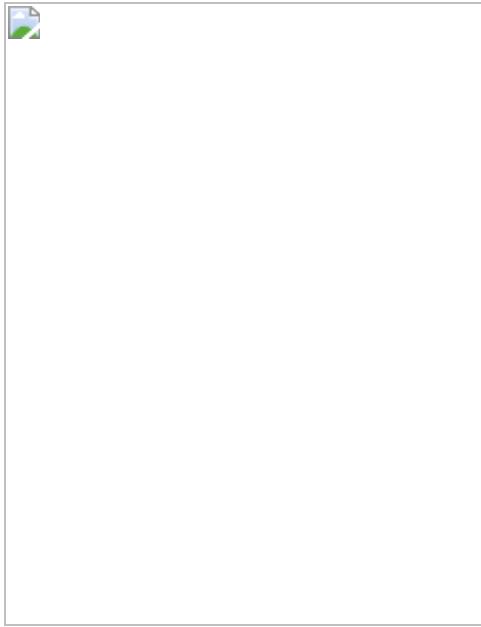
### **Additional access options:**

- [Log in](#)
- [Learn about institutional subscriptions](#)

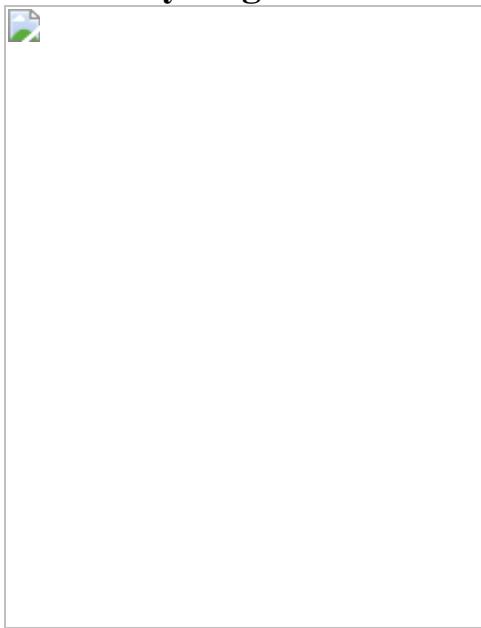
**Fig. 1: The structure and function of dolphin prestin homodimer in Cl<sup>-</sup>.**



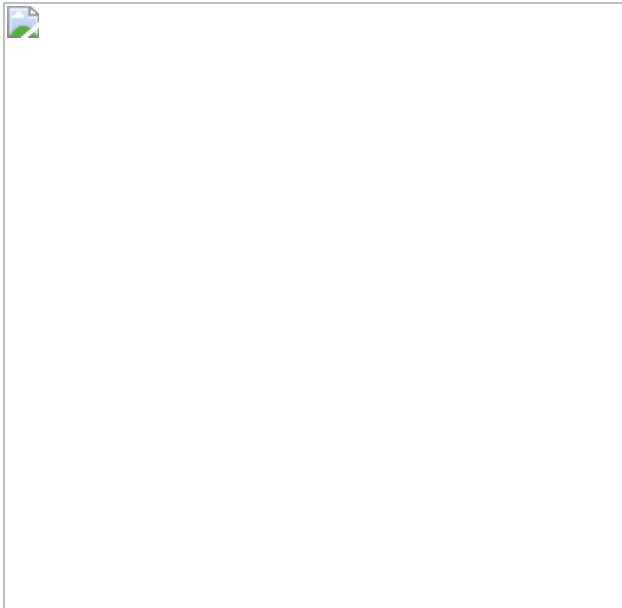
**Fig. 2: SO<sub>4</sub><sup>2-</sup> drives prestin towards the down and intermediate states at zero membrane potential.**



**Fig. 3: The structural basis of prestin inhibition by salicylate, and the evolutionary origins of electromotility.**



**Fig. 4: The structural basis of prestin's voltage sensitivity and somatic electromotility.**



## Data availability

The atomic structure coordinates have been deposited at the RCSB PDB under accession numbers [7S8X](#), [7S9A](#), [7S9B](#), [7S9C](#), [7S9D](#) and [7S9E](#); and the EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers [EMD-24928](#), [EMD-24930](#), [EMD-24931](#), [EMD-24932](#), [EMD-24933](#) and [EMD-24934](#). All materials generated during the current study are available from the corresponding author under a materials transfer agreement with The University of Chicago.

## References

1. 1.

Zheng, J. et al. Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**, 149–155 (2000).

2. 2.

Liu, X. Z. et al. Prestin, a cochlear motor protein, is defective in non-syndromic hearing loss. *Hum. Mol. Genet.* **12**, 1155–1162 (2003).

3. 3.

Liberman, M. C. et al. Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature* **419**, 300–304 (2002).

4. 4.

Dallos, P. et al. Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification. *Neuron* **58**, 333–339 (2008).

5. 5.

Cheatham, M., Huynh, K., Gao, J., Zuo, J. & Dallos, P. Cochlear function in Prestin knockout mice. *J. Physiol.* **560**, 821–830 (2004).

6. 6.

Masterton, B., Heffner, H. & Ravizza, R. The evolution of human hearing. *J. Acoust. Soc. Am.* **45**, 966–985 (1969).

7. 7.

Heffner, H. & Masterton, B. Hearing in glires: domestic rabbit, cotton rat, feral house mouse, and kangaroo rat. *J. Acoust. Soc. Am.* **68**, 1584–1599 (1980).

8. 8.

Fettiplace, R. Diverse mechanisms of sound frequency discrimination in the vertebrate cochlea. *Trends Neurosci.* **43**, 88–102 (2020).

9. 9.

Ashmore, J. A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol.* **388**, 323–347 (1987).

10. 10.

Brownell, W. E., Bader, C. R., Bertrand, D. & De Ribaupierre, Y. Evoked mechanical responses of isolated cochlear outer hair cells. *Science* **227**, 194–196 (1985).

11. 11.

Kachar, B., Brownell, W. E., Altschuler, R. & Fex, J. Electrokinetic shape changes of cochlear outer hair cells. *Nature* **322**, 365–368 (1986).

12. 12.

He, D. Z. et al. Changes in plasma membrane structure and electromotile properties in prestin deficient outer hair cells. *Cytoskeleton* **67**, 43–55 (2010).

13. 13.

Gorbunov, D. et al. Molecular architecture and the structural basis for anion interaction in prestin and SLC26 transporters. *Nat. Commun.* **5**, 3622 (2014).

14. 14.

Hallworth, R. & Nichols, M. G. Prestin in HEK cells is an obligate tetramer. *J. Neurophysiol.* **107**, 5–11 (2012).

15. 15.

Zheng, J. et al. Analysis of the oligomeric structure of the motor protein prestin. *J. Biol. Chem.* **281**, 19916–19924 (2006).

16. 16.

Navaratnam, D., Bai, J.-P., Samaranayake, H. & Santos-Sacchi, J. N-terminal-mediated homomultimerization of prestin, the outer hair cell motor protein. *Biophys. J.* **89**, 3345–3352 (2005).

17. 17.

He, D. Z., Lovas, S., Ai, Y., Li, Y. & Beisel, K. W. Prestin at year 14: progress and prospect. *Hear. Res.* **311**, 25–35 (2014).

18. 18.

Liu, Z., Qi, F.-Y., Xu, D.-M., Zhou, X. & Shi, P. Genomic and functional evidence reveals molecular insights into the origin of echolocation in whales. *Sci. Adv.* **4**, eaat8821 (2018).

19. 19.

Walter, J. D., Sawicka, M. & Dutzler, R. Cryo-EM structures and functional characterization of murine Slc26a9 reveal mechanism of uncoupled chloride transport. *eLife* **8**, e46986 (2019).

20. 20.

Chi, X. et al. Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. *Cell Discov.* **6**, 55 (2020).

21. 21.

Rybalchenko, V. & Santos-Sacchi, J. Anion control of voltage sensing by the motor protein prestin in outer hair cells. *Biophys. J.* **95**, 4439–4447 (2008).

22. 22.

Tan, X. et al. From zebrafish to mammal: functional evolution of prestin, the motor protein of cochlear outer hair cells. *J. Neurophysiol.* **105**, 36–44 (2011).

23. 23.

Santos-Sacchi, J., Song, L., Zheng, J. & Nuttall, A. L. Control of mammalian cochlear amplification by chloride anions. *J. Neurosci.* **26**, 3992–3998 (2006).

24. 24.

Rybalchenko, V. & Santos-Sacchi, J. Cl<sup>-</sup> flux through a non-selective, stretch-sensitive conductance influences the outer hair cell motor of the guinea-pig. *J. Physiol.* **547**, 873–891 (2003).

25. 25.

Homma, K., Duan, C., Zheng, J., Cheatham, M. A. & Dallos, P. The V499G/Y501H mutation impairs fast motor kinetics of prestin and has significance for defining functional independence of individual prestin subunits. *J. Biol. Chem.* **288**, 2452–2463 (2013).

26. 26.

Homma, K. & Dallos, P. Evidence that prestin has at least two voltage-dependent steps. *J. Biol. Chem.* **286**, 2297–2307 (2011).

27. 27.

Iwasa, K. A two-state piezoelectric model for outer hair cell motility. *Biophys. J.* **81**, 2495–2506 (2001).

28. 28.

Ludwig, J. et al. Reciprocal electromechanical properties of rat prestin: the motor molecule from rat outer hair cells. *Proc. Natl Acad. Sci. USA* **98**, 4178–4183 (2001).

29. 29.

Cazals, Y. Auditory sensori-neural alterations induced by salicylate. *Prog. Neurobiol.* **62**, 583–631 (2000).

30. 30.

Chen, G.-D. et al. Salicylate-induced cochlear impairments, cortical hyperactivity and re-tuning, and tinnitus. *Hear. Res.* **295**, 100–113 (2013).

31. 31.

Oliver, D. et al. Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science* **292**, 2340–2343 (2001).

32. 32.

Kakehata, S. & Santos-Sacchi, J. Effects of salicylate and lanthanides on outer hair cell motility and associated gating charge. *J. Neurosci.* **16**, 4881–4889 (1996).

33. 33.

Schaechinger, T. J. & Oliver, D. Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. *Proc. Natl Acad. Sci. USA* **104**, 7693–7698 (2007).

34. 34.

Song, L. & Santos-Sacchi, J. Conformational state-dependent anion binding in prestin: evidence for allosteric modulation. *Biophys. J.* **98**, 371–376 (2010).

35. 35.

Bai, J.-P. et al. Current carried by the Slc26 family member prestin does not flow through the transporter pathway. *Sci. Rep.* **7**, 46619 (2017).

36. 36.

Bai, J.-P. et al. Prestin's anion transport and voltage-sensing capabilities are independent. *Biophys. J.* **96**, 3179–3186 (2009).

37. 37.

Jogini, V. & Roux, B. Dynamics of the Kv1.2 voltage-gated K<sup>+</sup> channel in a membrane environment. *Biophys. J.* **93**, 3070–3082 (2007).

38. 38.

Starace, D. M. & Bezanilla, F. A proton pore in a potassium channel voltage sensor reveals a focused electric field. *Nature* **427**, 548–553 (2004).

39. 39.

Bezanilla, F. How membrane proteins sense voltage. *Nat. Rev. Mol. Cell Biol.* **9**, 323–332 (2008).

40. 40.

Dong, X.-X., Ehrenstein, D. & Iwasa, K. Fluctuation of motor charge in the lateral membrane of the cochlear outer hair cell. *Biophys. J.* **79**, 1876–1882 (2000).

41. 41.

Dong, X.-X. & Iwasa, K. Tension sensitivity of prestin: comparison with the membrane motor in outer hair cells. *Biophys. J.* **86**, 1201–1208 (2004).

42. 42.

Beurg, M., Tan, X. & Fettiplace, R. A prestin motor in chicken auditory hair cells: active force generation in a nonmammalian species. *Neuron* **79**, 69–81 (2013).

43. 43.

Dallos, P. & He, D. Z. Two models of outer hair cell stiffness and motility. *J. Assoc. Res. Otolaryngol.* **1**, 283–291 (2000).

44. 44.

Izumi, C., Bird, J. E. & Iwasa, K. H. Membrane thickness sensitivity of prestin orthologs: the evolution of a piezoelectric protein. *Biophys. J.* **100**, 2614–2622 (2011).

45. 45.

Fang, J., Izumi, C. & Iwasa, K. H. Sensitivity of prestin-based membrane motor to membrane thickness. *Biophys. J.* **98**, 2831–2838 (2010).

46. 46.

Santos-Sacchi, J., Shen, W., Zheng, J. & Dallos, P. Effects of membrane potential and tension on prestin, the outer hair cell lateral membrane motor protein. *J. Physiol.* **531**, 661–666 (2001).

47. 47.

Kakehata, S. & Santos-Sacchi, J. Membrane tension directly shifts voltage dependence of outer hair cell motility and associated gating charge. *Biophys. J.* **68**, 2190–2197 (1995).

48. 48.

Ashmore, J. Cochlear outer hair cell motility. *Physiol. Rev.* **88**, 173–210 (2008).

49. 49.

He, D. Z., Evans, B. N. & Dallos, P. First appearance and development of electromotility in neonatal gerbil outer hair cells. *Hear. Res.* **78**, 77–90 (1994).

50. 50.

Dallos, P. The active cochlea. *J. Neurosci.* **12**, 4575–4585 (1992).

51. 51.

Santos-Sacchi, J. Asymmetry in voltage-dependent movements of isolated outer hair cells from the organ of Corti. *J. Neurosci.* **9**, 2954–2962 (1989).

52. 52.

Santos-Sacchi, J. & Dilger, J. Whole cell currents and mechanical responses of isolated outer hair cells. *Hear. Res.* **35**, 143–150 (1988).

53. 53.

Ge, J. et al. Molecular mechanism of prestin electromotive signal amplification. *Cell* **184**, 4669–4679 (2021).

54. 54.

Butan, C. et al. Single particle cryo-EM structure of the outer hair cell motor protein prestin. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.08.03.454998> (2021).

55. 55.

Santos-Sacchi, J. On the frequency limit and phase of outer hair cell motility: effects of the membrane filter. *J. Neurosci.* **12**, 1906–1916 (1992).

56. 56.

Holley, M. & Ashmore, J. F. On the mechanism of a high-frequency force generator in outer hair cells isolated from the guinea pig cochlea. *Proc. R. Soc. Lond. B Biol. Sci.* **232**, 413–429 (1988).

57. 57.

Kirchhofer, A. et al. Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).

58. 58.

Clark, M. D., Contreras, G. F., Shen, R. & Perozo, E. Electromechanical coupling in the hyperpolarization-activated K<sup>+</sup> channel KAT1. *Nature* **583**, 145–149 (2020).

59. 59.

Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

60. 60.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

61. 61.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

62. 62.

Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

63. 63.

Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).

64. 64.

Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

65. 65.

Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

66. 66.

Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).

67. 67.

Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).

68. 68.

Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

69. 69.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

70. 70.

Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).

71. 71.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

72. 72.

Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).

73. 73.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

74. 74.

Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

75. 75.

Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

76. 76.

Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

77. 77.

Lindau, M. & Neher, E. Patch-clamp techniques for time-resolved capacitance measurements in single cells. *Pflügers Arch.* **411**, 137–146 (1988).

78. 78.

Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).

79. 79.

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

80. 80.

Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

81. 81.

Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

82. 82.

Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

83. 83.

Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).

84. 84.

Morozenko, A. & Stuchebrukhov, A. Dowser++, a new method of hydrating protein structures. *Proteins Struct. Funct. Bioinform.* **84**, 1347–1357 (2016).

85. 85.

Shaw, D. E. et al. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **51**, 91–97 (2008).

86. 86.

Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

87. 87.

Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

88. 88.

MacKerell, Jr A. D., Feig, M. & Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *JACS* **126**, 698–699 (2004).

89. 89.

MacKerell, Jr A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

90. 90.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

91. 91.

Huang, L. & Roux, B. Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data. *J. Chem. Theory Comput.* **9**, 3543–3556 (2013).

92. 92.

Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

93. 93.

Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).

94. 94.

Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

95. 95.

Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).

96. 96.

Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* **122**, 054101 (2005).

97. 97.

Aksimentiev, A. & Schulten, K. Imaging  $\alpha$ -hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* **88**, 3745–3761 (2005).

98. 98.

Roux, B. The membrane potential and its representation by a constant electric field in computer simulations. *Biophys. J.* **95**, 4205–4216 (2008).

99. 99.

Castillo, J. P. et al. Mechanism of potassium ion uptake by the  $\text{Na}^+/\text{K}^+$ -ATPase. *Nat. Commun.* **6**, 7622 (2015).

100. 100.

Khalili-Araghi, F. et al. Calculation of the gating charge for the Kv1.2 voltage-activated potassium channel. *Biophys. J.* **98**, 2189–2198 (2010).

101. 101.

Morin, A. et al. Cutting edge: Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

## Acknowledgements

We thank M. Zhao for discussions and assistance with imaging; K. Poole, C. Cox, T. Ngo, O. Bavi, R. Hulse, C. Bassetto, Y. Nikolaev, Y. Krishnan, P. Bezanilla, H. Mchaourab, P. Gueorguieva, S. Zhong, M. Karami, P. Haller and F. Galan, and the members of the Perozo laboratory for exchanging ideas and comments on the manuscript; P. Shi for sharing the *Tursiops* prestin plasmid; J. Fuller, J. Austin II and T. Lavoie at the University of Chicago Advanced Electron Microscopy Facility for microscope maintenance and training; and U. Baxa and T. J. Edwards at NCEF for cryo-EM data collection. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through grant R01GM116961 from the National Institutes of Health. The Anton 2 machine at PSC was made available by D.E. Shaw Research. N.B. acknowledges the Biology of Inner Ear course (BIE2019) and Gordon Conference (Auditory System Gordon Research Conference) for inspiring him to study hearing and prestin. This work was supported by NIDCD grant R01 DC019833 to E.P. N.B. was the recipient of a Chicago Fellowship. M.D.C. was supported by F30MH116647 and T32GM007281. This research was in part supported by the National Cancer Institute's National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research under contract HSSN261200800001E.

## Author information

### Author notes

1. Bharat G. Reddy

Present address: Rectify Pharmaceuticals, Cambridge, MA, USA

## Affiliations

1. Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA

Navid Bavi, Michael David Clark, Gustavo F. Contreras, Rong Shen, Bharat G. Reddy, Wieslawa Milewski & Eduardo Perozo

2. Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior, The University of Chicago, Chicago, IL, USA

Eduardo Perozo

## Contributions

N.B., M.D.C. and E.P. conceived the project. N.B. expressed and purified the protein. N.B. and M.D.C. prepared cryo-grids. N.B. and B.G.R. performed EM data collection. N.B., B.G.R. and M.D.C. processed the cryo-EM data and built and refined the atomic models. N.B. and G.F.C. performed and analysed the electrophysiological experiments. R.S. carried out MD simulations and electrostatic calculations. N.B. and W.M. carried out molecular cloning, mutagenesis and created all of the expression constructs. N.B. and W.M. managed cell culture. All of the authors analysed the data. N.B. and E.P. wrote the manuscript with input from all of the other authors.

## Corresponding author

Correspondence to [Eduardo Perozo](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Rachelle Gaudet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Function, biochemistry and structural features of electromotile prestin.

**a**, Electromotility analysis of HEK293 cells transfected with wild-type dolphin prestin compared to GFP-only transfected cell (Mock-GFP). The cellular displacement was normalized based on the cell largest diameter,  $d_0$  (Fig. 1b). The normalized electromotility was  $0.05\pm0.02$  ( $n = 6$ ) versus  $0.008\pm0.002$  ( $n = 5$ ) for wild-type prestin and Mock-GFP, respectively. These values were measured at the depolarizing voltage step changing from +120 mV to -120 mV (mean  $\pm$  SD,  $n$  is the number of independent cells. One-sided Student t-test, unpaired,  $P=0.005$ ). **b**, Size-exclusion chromatography (SEC) curves of the full-length dolphin prestin purified in GDN, run on a Superose 6 column, in high Cl<sup>-</sup> (red) and SO<sub>4</sub><sup>2-</sup> (blue) based solution. The fractions indicated by black dotted lines in both represent purified proteins that were used for cryo-EM imaging. **c**, Purified dolphin prestin cryo-EM samples, run on a Stain-free SDS-PAGE gel, indicating size of ~80 kDa for the full-length prestin monomer (representation of  $n = 3$ ). **d**, Topology of dolphin prestin. Different domains are indicated by color; the gate domain is colored in blue, the core domain in red and the C- and N-termini as well as the STAS domain in grey. The transmembrane helices are numbered from 1 to 14. The N- and C-termini as well as the STAS domain are oriented towards the cytoplasm.

**Extended Data Fig. 2 Flow chart for the cryo-EM data processing and structure determination of the dolphin prestin in high Cl<sup>-</sup> condition.**

**a**, The final reconstruction has a nominal resolution of 3.3 Å (at FSC=0.143). The yellow scale bar on the micrograph represents 200 Å. All the images in this figure were created in UCSF ChimeraX.

**Extended Data Fig. 3 Structure of prestin in high Cl<sup>-</sup> and comparison with the Intermediate.**

**a**, Comparison between prestin (high Cl<sup>-</sup>) (blue and Red) and SLC26A9 Intermediate state (6RTF, grey). The structures are aligned based on residues 460 to 505 of one subunit (TM13-TM14, dotted box). ChimeraX was used for illustration. **b**, Electrostatic potential and surface charge distribution of SLC26A9 intermediate state<sup>19</sup> compared with that of prestin in high Cl<sup>-</sup> panel **c**. The electrostatic charge distribution ranges from -5 to 5 kT from negative to positive charge. ChimeraX was used for illustration.

**Extended Data Fig. 4 Flow chart for the cryo-EM data processing and structure determination of the dolphin prestin in SO<sub>4</sub><sup>2-</sup>.**

**a, b** Cryo-EM data processing and structure determination of the dolphin prestin in Down I (SO<sub>4</sub><sup>2-</sup>) and Down II (SO<sub>4</sub><sup>2-</sup>) states. A was obtained from Dataset I, which was combined with Class B from Dataset II. The final reconstruction yielded two structures, Down I (SO<sub>4</sub><sup>2-</sup>) and Down II (SO<sub>4</sub><sup>2-</sup>), which have nominal resolutions of 4.2 and 6.7 Å, respectively (at FSC=0.143). See Supplementary Figure 5 for the steps on how Class A and B were further processed. Evidence of both states was found in dataset II, however merging of datasets was required to improve resolution of states. **c**, Flow chart for the cryo-EM data processing and structure determination of the dolphin prestin in the Intermediate state (SO<sub>4</sub><sup>2-</sup>) (See [Methods](#) for

details). The final reconstruction has a nominal resolution of 4.6 Å (at FSC=0.143). UCSF ChimeraX was for illustration of all the structures. The yellow scale bar on all the micrographs represents 200 Å.

### **Extended Data Fig. 5 Prestin's cross-sectional area changes upon transition from Down to Up states.**

**a**, Upon the transition from Down to Up state and the movement of the anion-binding site, the most obvious changes are seen in the peripheral helices TM5b, TM6-TM7, and TM8. **b**, MD simulation of prestin in Up state is compared with the Inhibited II state ( $\text{Cl}^-$  and Salicylate) equilibrated in POPC lipid bilayers. The cross-sectional area of outer and inner monolayers with mapped leaflet coordinate in the Z direction (across the membrane thickness) using all-atom molecular dynamics simulations (1 $\mu\text{s}$ ).  $\Delta z$  shows movement of the phosphate group of the lipids in the Z (thickness) direction. The comparison was made between Up ( $\text{Cl}^-$ ) and Inhibited II ( $\text{SO}_4^{2-}$ ) states. The largest difference was observed at the location of the TM6 helix. **c**, Cross-sectional area calculations of the transmembrane domain of SLC26A9(12) along the hydrophobic thickness using CHARMM-membrane builder. Cross-sectional area change of SLC26A9 from Inward-facing to Intermediate states (6RTC and 6RTF) per monomer<sup>19</sup>. Note that prior to area calculation, the spatial arrangements of all the structures with respect to the hydrocarbon core of the lipid bilayer were first adjusted using the PPM server(30). The structures were aligned based on residues 460 to 505 (TM13-TM14). **d**, Comparison of the change in the micelle morphology between two salicylate-inhibited structures Inhibited I ( $\text{Cl}^-$ ) and Inhibited II ( $\text{SO}_4^{2-}$  + Salicylate) states. The overlay of the two states shows drastic changes in the micelle thickness especially around TM6 region in addition to the overall changes in the micelle in-lane direction, both indicative of major structural rearrangements between the two states. ChimeraX was used for illustration.

### **Extended Data Fig. 6 Salicylate outcompetes $\text{SO}_4^{2-}$ in binding to anion-binding pocket.**

**a**, The NLC measurements of HEK293T cells transfected with dolphin prestin in  $\text{SO}_4^{2-}$  ( $0.15 \pm 0.06$ ;  $n = 6$ ). The NLC of these cells were completely abrogated ( $0.01 \pm 0.01$ ) by 10 mM Na-Salicylate (mean  $\pm$  s.e.m.;  $n$ , is the number of independent cells. One-sided student's t-test, unpaired,  $P=0.01$ ) **b**, Density of Salicylate (orange) in the anion-binding site (blue) was resolved in the Inhibited II ( $\text{SO}_4^{2-}$ ) state of dolphin prestin. **c**, Sequence alignment of prestin and close SLC transporters across different species. Residues forming the anion-binding site are largely conserved (e.g. Q97, F101, F137). Putative voltage-sensing residue R399 in dolphin prestin is replaced by a valine in murine SLC26A9. Clustal Omega was used for the sequence alignments. ChimeraX was used for illustration.

**Extended Data Fig. 7 Flow chart for the cryo-EM data processing and structure determination of the dolphin prestin in the Salicylate-Inhibited states.**

Flow chart of the dolphin prestin in the **a**, Inhibited I state ( $\text{Cl}^- +$  Salicylate) and **b**, ( $\text{SO}_4^{2-} +$  Salicylate) The final reconstructions have a nominal resolution of 3.8 Å and 3.7 Å, respectively (at FSC=0.143). All the images in this figure were created in UCSF ChimeraX. The yellow scale bar on all the micrographs represents 200 Å.

**Extended Data Fig. 8 Electrostatic calculations and charge transfer of prestin across the membrane.**

**a**, Mutation of the key residues in the anion binding pocket either completely abolishes the NLC (R399Q) or right shifts the  $V_{1/2}$  by more than 80 mV (F101Y) to around  $+25 \pm 5$  mV (mean  $\pm$  s.e.m.;  $n$ , is the number of independent cells. One-sided Student t-test,  $P=0.001$ ); a similar effect has been observed in other prestin homologues using patch-clamp electrophysiology (51). **b**, Snapshots from the MD trajectories of the systems, and calculation of the electrostatic potential across the membrane at two states, the Down I state (with  $\text{SO}_4^{2-}$  in the left cavity, and without  $\text{SO}_4^{2-}$  in the right cavity) versus Up (with  $\text{Cl}^-$  in the left cavity and without any  $\text{Cl}^-$  in the right cavity). The  $x$ - $z$  plane is crossing the two central anion-

binding sites. In both models, the positive field is mainly focused around the transmembrane mid-plane and around the anion-binding site, creating an attractive (blue) field for the binding of the anion. However, in the Up state the field is more positive around the mid-plane compared to the corresponding region in the Intermediate state. In both cases, the presence of the anion only partially neutralizes (~35%) the positive field around the bilayer mid-plane. Note that the actual size of the simulation box is larger than what is illustrated here (see [Methods](#)). **c**, Averaged 1-D fraction of membrane potential in the z direction along the two central binding sites (shown as dashed blue lines in panel A with the central binding sites highlighted using the red cross symbols). The 1-D and 2-D maps were directly extracted from the ensemble averaged 3-D fraction of membrane potential map. The location of the phosphate atoms of the outer and inner lipid leaflets along the z axis was highlighted with dashed gray lines). **d**, Displacement of charge for prestin in the Up and Down I conformations at different transmembrane potentials. The gating charge between the two states is  $0.38 + 0.25$  e calculated as the offset constant between the linear fits. ( $n = 3$ ; data are mean  $\pm$  SD; One-sided Student's t-test;  $P=0.05$ ). **e**, R399 in both monomers have been mutated to Q, S and E in different systems to see the contribution of R399 residue to the positive charge at the bilayer mid-plane using electrostatic calculations. R399 mutation to polar residues shows that R399 has almost ~40% contribution the positive charge of the field at the bilayer mid-plane. The remainder likely comes from the TM3-TM10 helical dipole and other positive charges in this area.

### [Extended Data Fig. 9 Whole cell patch-clamp electrophysiology of the mutations of different glycine residues along the TM6 helix.](#)

All the individual data points, that has been averaged in Fig. [3f](#), has been presented here. Compared to wild-type prestin, mutation of evolutionary conserved glycine residues, **a**, G274 and G275 and **b**, G263, G265 and G270 largely affects the NLC.

### **Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics**

# **Supplementary information**

## **Supplementary Information**

This file contains further supportive results for the findings in this study, including Supplementary Figs. 1–10 (which include the uncropped gel source data).

## **Reporting Summary**

## **Peer Review File**

## **Supplementary Video 1**

Electromotility measurements of HEK293 cells transfected with dolphin prestin using whole-cell patch-clamp electrophysiology. To evoke prestin-mediated electromotility, the membrane potential was held at  $-70$  mV;  $10$  mV increase-in-amplitude voltage steps were applied up to the final steps, which was from  $+150$  mV to  $-140$  mV (Fig. 1b). The magenta square indicates the area that was chosen in our custom-written code to track the cellular displacements.

## **Supplementary Video 2**

Structural changes from the expanded (down I) to the compact (up) conformation as a linear interpolation. The side front and top views of the dimer have been shown in one single frame. The anion-binding site is highlighted in red and Arg399 is shown in stick representation and the backbone has been coloured yellow. The videos were made in UCSF ChimeraX.

# **Rights and permissions**

## **Reprints and Permissions**

## About this article

### Cite this article

Bavi, N., Clark, M.D., Contreras, G.F. *et al.* The conformational cycle of prestin underlies outer-hair cell electromotility. *Nature* **600**, 553–558 (2021). <https://doi.org/10.1038/s41586-021-04152-4>

- Received: 17 August 2021
- Accepted: 15 October 2021
- Published: 25 October 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04152-4>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04152-4>

- Matters Arising
- [Published: 15 December 2021](#)

# On the liquid–liquid phase transition of dense hydrogen

- [Valentin V. Karasiev](#) ORCID: [orcid.org/0000-0003-3445-6797](https://orcid.org/0000-0003-3445-6797)<sup>1</sup>,
- [Joshua Hinz](#)<sup>1</sup>,
- [S. X. Hu](#) ORCID: [orcid.org/0000-0003-2465-3818](https://orcid.org/0000-0003-2465-3818)<sup>1</sup> &
- [S. B. Trickey](#) ORCID: [orcid.org/0000-0001-9224-6304](https://orcid.org/0000-0001-9224-6304)<sup>2</sup>

*Nature* volume **600**, pages E12–E14 (2021)

- 747 Accesses
- 27 Altmetric
- [Metrics details](#)

## Subjects

- [Giant planets](#)
- [Phase transitions and critical phenomena](#)

[Matters Arising](#) to this article was published on 15 December 2021

The [Original Article](#) was published on 09 September 2020

arising from B. Cheng et al. *Nature* <https://doi.org/10.1038/s41586-020-2677-y> (2020)

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

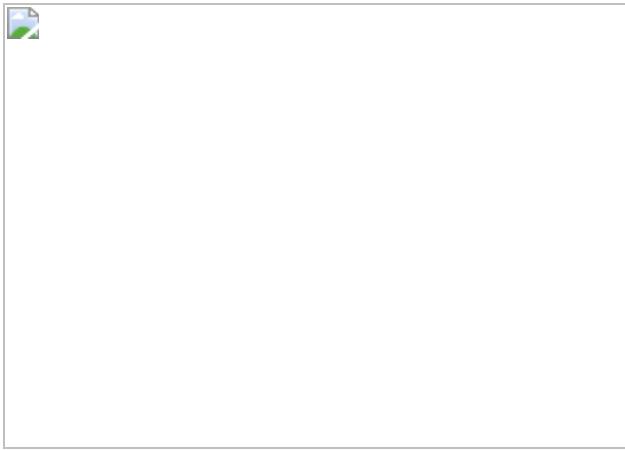
### Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Comparison of MD results from the PBE exchange-correlation-based MLP and ab initio DFT-MD *NPT* simulations.**



**Fig. 2: The LLPT boundary from the present large-scale DFT-MD (DFT/PBE) simulations compared to MLP (MLP/PBE)  $C_P^{\max}$  and  $\rho^{\max}$  curves.**



## Data availability

The data that support the findings shown in the figures are available from the corresponding author upon reasonable request.

## References

1. 1.

Gregoryanz, E. et al. Everything you always wanted to know about metallic hydrogen but were afraid to ask. *Matter Radiat. Extrem.* **5**, 038101 (2020).

2. 2.

Hinz, J. et al. Fully consistent density functional theory determination of the insulator-metal transition boundary in warm dense hydrogen. *Phys. Rev. Res.* **2**, 032065(R) (2020).

3. 3.

Rillo, G., Morales, M. A., Ceperley, D. M. & Pierleoni, C. Optical properties of high-pressure fluid hydrogen across molecular dissociation. *Proc. Natl Acad. Sci.* **116**, 9770–9774 (2019).

4. 4.

Lu, B., Kang, D., Wang, D., Gao, T. & Dai, J. Towards the same line of liquid–liquid phase transition of dense hydrogen from various theoretical predictions. *Chin. Phys. Lett.* **36**, 103102 (2019).

5. 5.

Pierleoni, C., Morales, M. A., Rillo, G., Holzmann, M. & Ceperley, D. M. Liquid–liquid phase transition in hydrogen by coupled electron–ion Monte Carlo simulations. *Proc. Natl Acad. Sci.* **113**, 4953–4957, (2016).

6. 6.

Cheng, B., Mazzola, G., Pickard, C. J. & Ceriotti, M. Evidence for supercritical behaviour of high-pressure liquid hydrogen. *Nature* **585**, 217–220 (2020).

7. 7.

Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996);

erratum **78**, 1396 (1997).

8. 8.

Baldereschi, A. Mean-value point in the Brillouin zone. *Phys. Rev. B* **7**, 5212–5215 (1973).

9. 9.

Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

10. 10.

Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

11. 11.

Kapil, V. et al. I-PI 2.0: a universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **236**, 214–223 (2019).

12. 12.

Giannozzi, P. et al. Advanced capabilities for materials modelling with quantum espresso. *J. Phys. Condens. Matter* **29**, 465901 (2017).

13. 13.

Lorenzen, W., Holst, B. & Redmer, R. First-order liquid-liquid phase transition in dense hydrogen. *Phys. Rev. B* **82**, 195107 (2010).

14. 14.

Zha, C. S., Liu, H., Tse, J. S. & Hemley, R. J. Melting and high  $P-T$  transitions of hydrogen up to 300 GPa. *Phys. Rev. Lett.* **119**, 075302 (2017).

15. 15.

Geng, H. Y., Wu, Q., Marqués, M. & Ackland, G. J. Thermodynamic anomalies and three distinct liquid-liquid transitions in warm dense liquid hydrogen. *Phys. Rev. B* **100**, 134109 (2019).

## Acknowledgements

This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S.

Government or any agency thereof. V.V.K., J.H. and S.X.H. were supported by the Department of Energy National Nuclear Security Administration Award Number DE-NA0003856 and US National Science Foundation PHY Grant No. 1802964. Partial funding for S.X.H. was provided by the NSF Physics Frontier Center Award PHY-2020249. S.B.T. was supported by Department of Energy Grant DE-SC0002139. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Part of the computations were performed on the Laboratory for Laser Energetics HPC systems.

## Author information

### Affiliations

1. Laboratory for Laser Energetics, University of Rochester, Rochester, NY, USA

Valentin V. Karasiev, Joshua Hinz & S. X. Hu

2. Quantum Theory Project, Department of Physics, University of Florida, Gainesville, FL, USA

S. B. Trickey

## Contributions

V.V.K. conceived the project and designed the study. V.V.K. and J.H. performed the DFT-MD simulations and postprocessed the data. V.V.K. wrote the initial manuscript with inputs from S.X.H. S.B.T. revised the conception and scope. V.V.K. and S.B.T rewrote the manuscript. All authors discussed the results and revised the paper extensively.

## Corresponding author

Correspondence to [Valentin V. Karasiev](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

## Supplementary Information

This file contains Supplementary Information, Figs. 1–3, refs. 1–7.

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Karasiev, V.V., Hinz, J., Hu, S.X. *et al.* On the liquid–liquid phase transition of dense hydrogen. *Nature* **600**, E12–E14 (2021).  
<https://doi.org/10.1038/s41586-021-04078-x>

- Received: 26 December 2020
- Accepted: 30 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04078-x>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04078-x>

| [Section menu](#) | [Main menu](#) |

- Matters Arising
- [Published: 15 December 2021](#)

# Reply to: On the liquid–liquid phase transition of dense hydrogen

- [Bingqing Cheng](#) ORCID: [orcid.org/0000-0002-3584-9632<sup>1</sup>](https://orcid.org/0000-0002-3584-9632),
- [Guglielmo Mazzola](#) ORCID: [orcid.org/0000-0002-8982-5169<sup>2</sup>](https://orcid.org/0000-0002-8982-5169),
- [Chris J. Pickard](#) ORCID: [orcid.org/0000-0002-9684-5432<sup>3</sup>](https://orcid.org/0000-0002-9684-5432) &
- [Michele Ceriotti](#) ORCID: [orcid.org/0000-0003-2571-2832<sup>4</sup>](https://orcid.org/0000-0003-2571-2832)

*Nature* volume 600, pages E15–E16 (2021)

- 571 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Astronomy and planetary science](#)
- [Condensed-matter physics](#)

The [Original Article](#) was published on 15 December 2021

replying to V. V. Karasiev et al. *Nature* <https://doi.org/10.1038/s41586-021-04078-x> (2021)

This is a preview of subscription content

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

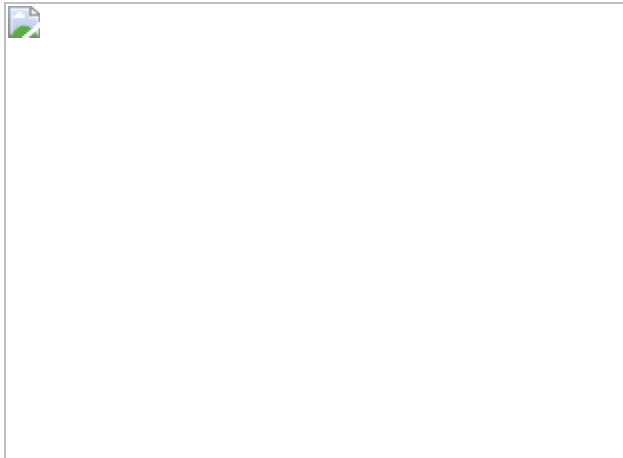
[Rent or Buy](#)

All prices are NET prices.

### Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: A comparison of the potential energies and pressures predicted by the MLP and ref. <sup>8</sup> using PBE DFT.**



## Data availability

The data behind ref. <sup>1</sup> can be found at GitHub (<https://github.com/BingqingCheng/MLP-highP-H>).

## References

1. 1.

Cheng, B., Mazzola, G., Pickard, C. J. & Ceriotti, M., Evidence for supercritical behaviour of high-pressure liquid hydrogen *Nature* **585**, 217–220 (2020).

2. 2.

Karasiev, V. V., Hinz, J., Hu, S. X. & Trickey, S. B. On the liquid–liquid phase transition of dense hydrogen. *Nature*, <https://doi.org/10.1038/s41586-021-04078-x> (2021).

3. 3.

Cheng, B., Mazzola, G. & Ceriotti, M. Evidence for supercritical behavior of high-pressure liquid. Preprint at <https://arxiv.org/abs/1906.03341> (2019).

4. 4.

Goncharov, A. F., Chuvashova, I., Ji, C. & Mao, H. Intermolecular coupling and fluxional behavior of hydrogen in phase IV. *Proc. Natl Acad. Sci.* **116**, 25512–25515 (2019).

5. 5.

Liu, H., & Ma, Y. Proton or deuteron transfer in phase IV of solid hydrogen and deuterium. *Phys. Rev. Lett.* **110**, 025903 (2013).

6. 6.

Lorenzen, W., Holst, B. & Redmer, R. First-order liquid-liquid phase transition in dense hydrogen. *Phys. Rev. B* **82**, 195107 (2010).

7. 7.

Zha, C., Liu, H., John, S. T. & Hemley, R. J. Melting and high  $P-T$  transitions of hydrogen up to 300 GPa. *Phys. Rev. Lett.* **119**, 075302 (2017).

8. 8.

Geng, H. Y., Wu, Q., Marques, M. & Ackland, G. J. Thermodynamic anomalies and three distinct liquid-liquid transitions in warm dense liquid hydrogen. *Phys. Rev. B* **100**, 134109 (2019).

## Acknowledgements

We thank G. Ackland, H. Geng and R. Redmer, who shared their ab initio MD trajectories for us to benchmark the MLP during the review process of our original paper.

## Author information

### Affiliations

1. Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

Bingqing Cheng

2. IBM Quantum, IBM Research – Zurich, Ruschlikon, Switzerland

Guglielmo Mazzola

3. Department of Materials Science & Metallurgy, University of Cambridge, Cambridge, UK

Chris J. Pickard

4. Laboratory of Computational Science and Modeling, Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Michele Ceriotti

## Contributions

B.C., G.M., C.J.P. and M.C. performed research and wrote the paper.

## Corresponding author

Correspondence to [Bingqing Cheng](#).

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Cheng, B., Mazzola, G., Pickard, C.J. *et al.* Reply to: On the liquid–liquid phase transition of dense hydrogen. *Nature* **600**, E15–E16 (2021).  
<https://doi.org/10.1038/s41586-021-04079-w>

- Received: 12 February 2021
- Accepted: 30 September 2021
- Published: 15 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04079-w>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

# Evidence for supercritical behaviour of high-pressure liquid hydrogen

- Bingqing Cheng
- Guglielmo Mazzola
- Michele Ceriotti

Article 09 Sept 2020

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04079-w>

| [Section menu](#) | [Main menu](#) |

# Amendments & Corrections

- [\*\*Author Correction: Estimating a social cost of carbon for global energy consumption\*\*](#) [ 29 November 2021]  
Author Correction •
- [\*\*Author Correction: Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring\*\*](#) [ 30 November 2021]  
Author Correction •
- [\*\*Publisher Correction: Single-photon nonlinearity at room temperature\*\*](#) [ 02 December 2021]  
Publisher Correction •

- Author Correction
- [Published: 29 November 2021](#)

# Author Correction: Estimating a social cost of carbon for global energy consumption

- [Ashwin Rode](#) [ORCID: orcid.org/0000-0003-1369-4331<sup>1</sup>](#),
- [Tamma Carleton](#) [ORCID: orcid.org/0000-0002-5518-0550<sup>2,3</sup>](#),
- [Michael Delgado<sup>4</sup>](#),
- [Michael Greenstone](#) [ORCID: orcid.org/0000-0002-2364-2810<sup>3,5</sup>](#),
- [Trevor Houser<sup>4</sup>](#),
- [Solomon Hsiang](#) [ORCID: orcid.org/0000-0002-2074-0829<sup>3,6</sup>](#),
- [Andrew Hultgren](#) [ORCID: orcid.org/0000-0002-6720-6252<sup>1,5</sup>](#),
- [Amir Jina](#) [ORCID: orcid.org/0000-0003-3446-7883<sup>3,7</sup>](#),
- [Robert E. Kopp](#) [ORCID: orcid.org/0000-0003-4016-9428<sup>8,9</sup>](#),
- [Kelly E. McCusker<sup>4</sup>](#),
- [Ishan Nath<sup>10</sup>](#),
- [James Rising](#) [ORCID: orcid.org/0000-0001-8514-4748<sup>11</sup>](#) &
- [Jiacan Yuan<sup>12,13,14,15</sup>](#)

[Nature](#) volume 600, page E17 (2021)

- 714 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Climate-change impacts](#)
- [Energy economics](#)
- [Environmental economics](#)

The [Original Article](#) was published on 13 October 2021

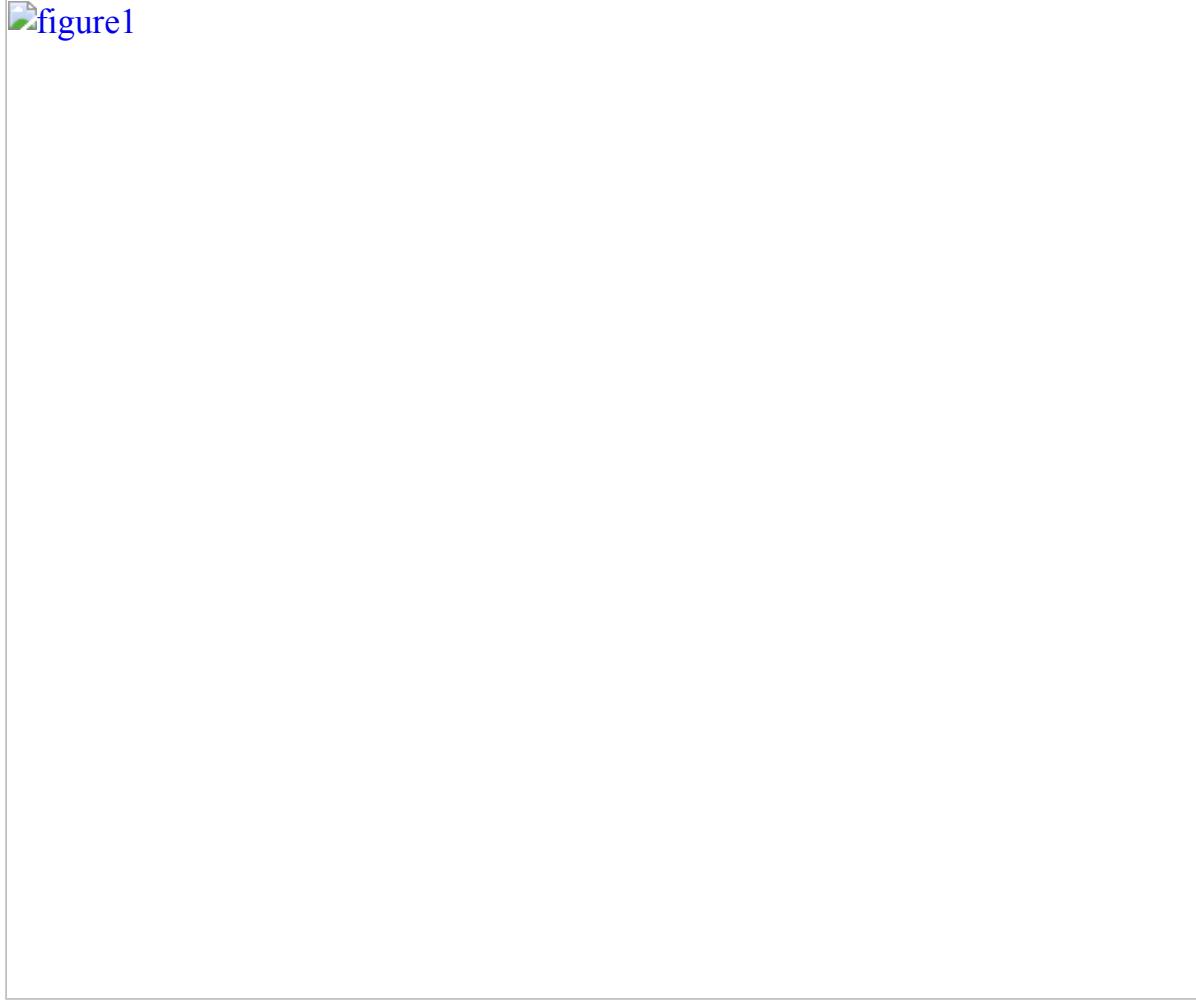
[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03883-8> Published online 13 October 2021

In the version of this Article initially published, there was an error in Fig. 2. In panel a, left-hand “Electricity” side, the gradient of colours shown in the map was inconsistent with the range in the colour scale bar. Original and corrected images are shown below in Fig. 1. The colour range in the map has been corrected in the online version of the article.

**Fig. 1: Original (incorrect) image (top) and corrected image (bottom) for Fig 2a.**

 figure1



## Author information

## Affiliations

1. Energy Policy Institute, University of Chicago, Chicago, IL, USA

Ashwin Rode & Andrew Hultgren

2. Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA, USA

Tamma Carleton

3. National Bureau of Economic Research, Cambridge, MA, USA

Tamma Carleton, Michael Greenstone, Solomon Hsiang & Amir Jina

4. Rhodium Group, New York, NY, USA

Michael Delgado, Trevor Houser & Kelly E. McCusker

5. Department of Economics, University of Chicago, Chicago, IL, USA

Michael Greenstone & Andrew Hultgren

6. Global Policy Laboratory, Goldman School of Public Policy, University of California, Berkeley, Berkeley, CA, USA

Solomon Hsiang

7. Harris School of Public Policy, University of Chicago, Chicago, IL, USA

Amir Jina

8. Department of Earth & Planetary Sciences, Rutgers University, New Brunswick, NJ, USA

Robert E. Kopp

9. Rutgers Institute of Earth, Ocean, and Atmospheric Sciences, Rutgers University, New Brunswick, NJ, USA

Robert E. Kopp

10. Department of Economics, Princeton University, Princeton, NJ, USA

Ishan Nath

11. School of Marine Science and Policy, University of Delaware, Newark, DE, USA

James Rising

12. Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai, China

Jiacan Yuan

13. Institute of Atmospheric Sciences, Fudan University, Shanghai, China

Jiacan Yuan

14. IRDR ICoE-RIG-WECEIPHE, Fudan University, Shanghai, China

Jiacan Yuan

15. Big Data Institute for Carbon Emission and Environmental Pollution, Fudan University, Shanghai, China

Jiacan Yuan

## Corresponding authors

Correspondence to [Ashwin Rode](#) or [Solomon Hsiang](#).

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Rode, A., Carleton, T., Delgado, M. *et al.* Author Correction: Estimating a social cost of carbon for global energy consumption. *Nature* **600**, E17 (2021).  
<https://doi.org/10.1038/s41586-021-04185-9>

- Published: 29 November 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04185-9>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04185-9>

| [Section menu](#) | [Main menu](#) |

- Author Correction
- [Published: 30 November 2021](#)

# Author Correction: Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring

- [Yu Shi](#)<sup>1</sup>,
- [Weina Gao](#)<sup>2,3</sup> na1,
- [Nikki K. Lytle](#)<sup>4,5</sup> na1,
- [Peiwu Huang](#)<sup>2,3,6</sup>,
- [Xiao Yuan](#)<sup>2,3</sup>,
- [Amanda M. Dann](#)<sup>7</sup>,
- [Maya Ridinger-Saison](#)<sup>8</sup> nAff24,
- [Kathleen E. DelGiorno](#)<sup>8</sup>,
- [Corina E. Antal](#)<sup>8</sup>,
- [Gaoyang Liang](#)<sup>8</sup>,
- [Annette R. Atkins](#)<sup>8</sup>,
- [Galina Erikson](#)<sup>9</sup>,
- [Huaiyu Sun](#)<sup>1</sup>,
- [Jill Meisenhelder](#)<sup>1</sup>,
- [Elena Terenziani](#)<sup>1</sup> nAff25,
- [Gyunghwi Woo](#)<sup>1</sup>,
- [Linjing Fang](#)<sup>10</sup>,
- [Thom P. Santisakultarm](#)<sup>10</sup>,
- [Uri Manor](#)<sup>10</sup>,
- [Ruilian Xu](#)<sup>11</sup>,
- [Carlos R. Becerra](#)<sup>12</sup>,

- [Erkut Borazanci](#)<sup>13,14</sup>,
- [Daniel D. Von Hoff](#)<sup>13,14</sup>,
- [Paul M. Grandgenett](#)<sup>15</sup>,
- [Michael A. Hollingsworth](#)<sup>15</sup>,
- [Mathias Leblanc](#)<sup>8</sup>,
- [Sarah E. Umetsu](#)<sup>16</sup>,
- [Eric A. Collisson](#)<sup>17</sup>,
- [Miriam Scadeng](#)<sup>18</sup>,
- [Andrew M. Lowy](#)<sup>19</sup>,
- [Timothy R. Donahue](#)<sup>7</sup>,
- [Tannishtha Reya](#)<sup>4,5,20</sup>,
- [Michael Downes](#)<sup>8</sup>,
- [Ronald M. Evans](#)<sup>8,21</sup>,
- [Geoffrey M. Wahl](#)<sup>8</sup>,
- [Tony Pawson](#)<sup>22,23 na2</sup>,
- [Ruijun Tian](#) <sup>2,3,22</sup> &
- [Tony Hunter](#)   [ORCID: orcid.org/0000-0002-7691-6993](#)<sup>1</sup>

*Nature* volume 600, page E18 (2021)

- 662 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Cancer metabolism](#)
- [Cancer stem cells](#)
- [Pancreatic cancer](#)
- [Targeted therapies](#)
- [Tumour biomarkers](#)

The [Original Article](#) was published on 17 April 2019

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1130-6> Published online 17 April 2019.

In the version of this Letter initially published, the Acknowledgements section omitted the following note of support from Geoffrey M. Wahl: “Work in the laboratory of G.M.W. was supported, in part, by the Cancer Center Core Grant (CA014195), National Institutes of Health/National Cancer Institute (R35 CA197687), National Institutes of Health Cancer Training Grant (T32 CA009370), the Isacoff Research Foundation Gastrointestinal (ICOIF) and the Freeberg Foundation.”.

## Author information

### Author notes

1. Maya Ridinger-Saison

Present address: Trovagene, San Diego, CA, USA

2. Elena Terenziani

Present address: Crown Bioscience San Diego, San Diego, CA, USA

3. These authors contributed equally: Weina Gao, Nikki K. Lytle

4. Deceased: Tony Pawson

## Affiliations

1. Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA

Yu Shi, Huaiyu Sun, Jill Meisenhelder, Elena Terenziani, Gyunghwi Woo & Tony Hunter

2. Department of Chemistry, Southern University of Science and Technology, Shenzhen, China

Weina Gao, Peiwu Huang, Xiao Yuan & Ruijun Tian

3. Shenzhen Key Laboratory of Cell Microenvironment, Southern University of Science and Technology, Shenzhen, China

Weina Gao, Peiwu Huang, Xiao Yuan & Ruijun Tian

4. Department of Pharmacology, University of California San Diego School of Medicine, La Jolla, CA, USA

Nikki K. Lytle & Tannishtha Reya

5. Sanford Consortium for Regenerative Medicine, La Jolla, CA, USA

Nikki K. Lytle & Tannishtha Reya

6. State Key Laboratory of Environmental and Biological Analysis, Department of Chemistry, Hong Kong Baptist University, Hong Kong, China

Peiwu Huang

7. Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

Amanda M. Dann & Timothy R. Donahue

8. Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA

Maya Ridinger-Saison, Kathleen E. DelGiorno, Corina E. Antal, Gaoyang Liang, Annette R. Atkins, Mathias Leblanc, Michael Downes, Ronald M. Evans & Geoffrey M. Wahl

9. Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA, USA

Galina Erikson

10. Waitt Advanced Biophotonics Center, Salk Institute for Biological Studies, La Jolla, CA, USA

Linjing Fang, Thom P. Santisakultarm & Uri Manor

11. Institute of Oncology, Shenzhen People's Hospital, Shenzhen, China

Ruilian Xu

12. Texas Oncology—Baylor University Medical Center, Dallas, TX, USA

Carlos R. Becerra

13. The Translational Genomics Research Institute, Scottsdale, AZ, USA

Erkut Borazanci & Daniel D. Von Hoff

14. HonorHealth, Scottsdale, AZ, USA

Erkut Borazanci & Daniel D. Von Hoff

15. Eppley Institute for Research in Cancer and Allied Diseases, Fred & Pamela Buffett Cancer Center, University of Nebraska Medical Center, Omaha, NE, USA

Paul M. Grandgenett & Michael A. Hollingsworth

16. Department of Pathology, University of California San Francisco, San Francisco, CA, USA

Sarah E. Umetsu

17. Hematology Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

Eric A. Collisson

18. Center for Functional MRI, Department of Radiology, University of California San Diego, La Jolla, CA, USA

Miriam Scadeng

19. Department of Surgery, Division of Surgical Oncology, University of California San Diego School of Medicine, La Jolla, CA, USA

Andrew M. Lowy

20. Moores Cancer Center, University of California San Diego School of Medicine, La Jolla, CA, USA

Tannishtha Reya

21. Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA, USA

Ronald M. Evans

22. Lunenfeld—Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada

Tony Pawson & Ruijun Tian

23. Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Tony Pawson

## Corresponding authors

Correspondence to [Yu Shi](#), [Ruijun Tian](#) or [Tony Hunter](#).

## Rights and permissions

## [Reprints and Permissions](#)

## About this article

### Cite this article

Shi, Y., Gao, W., Lytle, N.K. *et al.* Author Correction: Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* **600**, E18 (2021). <https://doi.org/10.1038/s41586-021-04176-w>

- Published: 30 November 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04176-w>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04176-w>

- Publisher Correction
- [Published: 02 December 2021](#)

# Publisher Correction: Single-photon nonlinearity at room temperature

- [Anton V. Zasedatelev](#)<sup>1,2</sup>,
- [Anton V. Baranikov](#)<sup>1,2</sup>,
- [Denis Sannikov](#)<sup>1,2</sup>,
- [Darius Urbonas](#) [ORCID: orcid.org/0000-0001-9919-9221](#)<sup>3</sup>,
- [Fabio Scafirimuto](#) [ORCID: orcid.org/0000-0002-9707-3526](#)<sup>3</sup>,
- [Vladislav Yu. Shishkov](#)<sup>1,2,4,5</sup>,
- [Evgeny S. Andrianov](#)<sup>1,2,4,5</sup>,
- [Yuri E. Lozovik](#)<sup>1,2,4,6,7</sup>,
- [Ullrich Scherf](#) [ORCID: orcid.org/0000-0001-8368-4919](#)<sup>8</sup>,
- [Thilo Stöferle](#) [ORCID: orcid.org/0000-0003-0612-7195](#)<sup>3</sup>,
- [Rainer F. Mahrt](#) [ORCID: orcid.org/0000-0002-9772-1490](#)<sup>3</sup> &
- [Pavlos G. Lagoudakis](#) [ORCID: orcid.org/0000-0002-3557-5299](#)<sup>1,2,9</sup>

[Nature](#) volume 600, page E19 (2021)

- 592 Accesses
- 2 Altmetric
- [Metrics details](#)

## Subjects

- [Bose–Einstein condensates](#)

- [Photonic devices](#)
- [Single photons and quantum effects](#)

The [Original Article](#) was published on 22 September 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03866-9>

Published online 22 September 2021

In the version of this Article originally published, there was an error in Fig. 1b. Specifically, the *x*-axis label mistakenly reported units as micrometres rather than microjoules. The corrected axis label now reads: “Pump fluence ( $\mu\text{J cm}^{-2}$ ).” The original Article has been corrected online.

## Author information

### Affiliations

1. Center for Photonics and Quantum Materials, Skolkovo Institute of Science and Technology, Moscow, Russia

Anton V. Zasedatelev, Anton V. Baranikov, Denis Sannikov, Vladislav Yu. Shishkov, Evgeny S. Andrianov, Yurii E. Lozovik & Pavlos G. Lagoudakis

2. Laboratories for Hybrid Photonics, Skolkovo Institute of Science and Technology, Moscow, Russia

Anton V. Zasedatelev, Anton V. Baranikov, Denis Sannikov, Vladislav Yu. Shishkov, Evgeny S. Andrianov, Yurii E. Lozovik & Pavlos G. Lagoudakis

3. IBM Research Europe - Zurich, Rüschlikon, Switzerland

Darius Urbonas, Fabio Scafirimuto, Thilo Stöferle & Rainer F. Mahrt

4. Dukhov Research Institute of Automatics (VNIIA), Moscow, Russia

Vladislav Yu. Shishkov, Evgeny S. Andrianov & Yurii E. Lozovik

5. Moscow Institute of Physics and Technology, Dolgoprudny, Russia

Vladislav Yu. Shishkov & Evgeny S. Andrianov

6. Institute for Spectroscopy RAS, Troitsk, Russia

Yurii E. Lozovik

7. Moscow Institute of Electronics and Mathematics, National Research University Higher School of Economics, Moscow, Russia

Yurii E. Lozovik

8. Macromolecular Chemistry Group and Institute for Polymer Technology, Bergische Universität Wuppertal, Wuppertal, Germany

Ullrich Scherf

9. Department of Physics and Astronomy, University of Southampton, Southampton, UK

Pavlos G. Lagoudakis

## Corresponding authors

Correspondence to [Anton V. Zasedatelev](#) or [Pavlos G. Lagoudakis](#).

## Rights and permissions

[Reprints and Permissions](#)

## About this article

## Cite this article

Zasedatelev, A.V., Baranikov, A.V., Sannikov, D. *et al.* Publisher Correction: Single-photon nonlinearity at room temperature. *Nature* **600**, E19 (2021). <https://doi.org/10.1038/s41586-021-04113-x>

- Published: 02 December 2021
- Issue Date: 16 December 2021
- DOI: <https://doi.org/10.1038/s41586-021-04113-x>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04113-x>

# Nature Outlook