

nature



DRIVING FORCE

AI algorithm outcompetes human champions in *Gran Turismo* racing game

Safety net

Planning for the day quantum computers crack encryption

Global health

Reboot biomedical R&D so everyone benefits from discovery

Joined-up thinking

Technique makes grafting possible for key food crops

Nature.2022.02.12

[Sat, 12 Feb 2022]

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

This Week

- **[Long COVID and kids: more research is urgently needed](#)** [08 February 2022]
Editorial • Like adults, children can experience long COVID, but few studies of the condition include young people. That has to change.
- **[Africa is bringing vaccine manufacturing home](#)** [09 February 2022]
Editorial • A major milestone was reached last week when scientists in South Africa reproduced Moderna's COVID-19 vaccine. COVID-19 patents must now be shared.
- **[Tracking COVID-19 infections: time for change](#)** [08 February 2022]
World View • We need better numbers if we are to manage the pandemic.
- **[‘Spines on posts’ hint at ancient devotion to the dead](#)** [02 February 2022]
Research Highlight • Indigenous people in what is now Peru partially re-assembled human skeletons, perhaps after desecration by European tomb robbers.
- **[Killer whales teach pals a new snack source: fishing lines](#)** [26 January 2022]
Research Highlight • Orcas in one area of the southern Indian Ocean now snatch some 180 tonnes of valuable Patagonian toothfish from fishing lines every year.
- **[Gas stoves help to cook the climate — even when off](#)** [04 February 2022]
Research Highlight • US kitchen appliances that burn natural gas emit much more methane than previously realized.
- **[Microbes that inflict deadly diarrhoea gain new powers](#)** [31 January 2022]
Research Highlight • Shigella bacteria, which kill more than 200,000 people every year, are evolving resistance to crucial antibiotics.
- **[Rare metal helps to turn sunlight into fuel, day and night](#)** [01 February 2022]
Research Highlight • Molecule that incorporates a light-sensitive ruthenium compound can store solar energy, then release it later to form hydrogen fuel.
- **[Huge cancer database yields genetic clues to metastasis](#)** [03 February 2022]

Research Highlight • Scientists sift through samples from more than 25,000 people to identify a host of mutations that could make tumours more aggressive.

- **How to protect unvaccinated children from COVID: vaccinate their parents** [03 February 2022]

Research Highlight • Data from more than 150,000 households show that jabs for adults drastically reduce young family members' risk of SARS-CoV-2 infection.

- EDITORIAL
- 08 February 2022

Long COVID and kids: more research is urgently needed

Like adults, children can experience long COVID, but few studies of the condition include young people. That has to change.



Venezuela is vaccinating its under-11s, a move that could help to reduce long COVID in children. Credit: Pedro Rances Mattey/Anadolu Agency/Getty

Long COVID is a term devised by patients to describe the lingering symptoms they experience well after an initial bout of COVID-19. The symptoms vary widely, but some of the most common are fatigue, shortness of breath, cognitive dysfunction (also called brain fog) and post-exertional malaise, in which even minor physical activity leads to lasting exhaustion. Between one-fifth and one-third of those with long COVID remain ill at least 12 weeks after a diagnosis of COVID-19, and a significant number continue to experience symptoms many months later. Many want the condition to be considered a disability.

Two years into the pandemic, there have been some 400 million confirmed cases of COVID-19 worldwide. Many more have probably gone undocumented. On the basis of the prevalence reported so far, there could be somewhere in the region of 100 million people now living with long COVID.

What little is known about long COVID in children and teenagers suggests that it can be just as disabling as it is in adults. However, there are many fewer studies in teens than in adults — and even fewer in children under the age of 11. This latter group is seeing a surge of COVID-19 infections: in many countries, children are not being vaccinated. More COVID-19 in kids will lead both to more long-COVID cases and to the spread of disease among vulnerable populations. It's time for younger people to be included in more studies of the condition, including trials of potential treatments. The UK support group Long Covid Kids says that reports of long COVID in children and teenagers are disbelieved by medical professionals. That, too, needs to change.

In adults, multiple studies have assessed the prevalence of long COVID, so there are now relatively good data showing that long COVID is alarmingly common. For example, a meta-analysis published last December ([F. Ceban](#) [Brain Behav. Immun. 101, 93–135; 2021](#)) pulled together 81 long-COVID studies published up to last June and examined how people were doing 12 or more weeks after being diagnosed with COVID-19 — 32% reported that they were still experiencing fatigue and 22% reported cognitive impairment. The bulk of these studies were done in high-income countries, where populations are older, and include few data on children and teenagers. But they do provide a sense of the scale of the problem.

Many of the studies of long COVID globally that do include children were done in hospitalized people. But some of the best evidence on long COVID in younger people comes from Children & Young People with Long Covid (CLoCk), a study by researchers at the University College London Great Ormond Street Institute of Child Health. The study, published this week (T. Stephenson *et al. Lancet Child Adolesc. Health* <https://doi.org/hf79>; 2022), recruited 6,804 11- to 17-year-olds in the United Kingdom in early 2021. About one-half had positive PCR tests for COVID-19; the other half were negative and served as controls. Three months after being tested, both groups completed a questionnaire asking what symptoms they were experiencing. Both reported some symptoms, but those who had tested positive were more likely to have long-COVID symptoms than were those with a negative test result — and were almost twice as likely to report three or more symptoms.

The CLoCk study suggests that, in the United Kingdom alone, tens of thousands of children and young people might have long COVID. This is in line with an estimate from the UK Office for National Statistics (see go.nature.com/3j7wx7t) that 44,000 2- to 11-year-olds in the country have long COVID, as do 73,000 12- to 16-year-olds. These figures are subject to uncertainties, but do still demonstrate that younger people are developing long COVID in significant numbers. It is irresponsible of governments to allow the virus to spread in this age group, especially in countries where the majority of children are unvaccinated. Past disease outbreaks often led to lasting symptoms, such as post-polio syndrome, and COVID-19 is clearly no different.

The few studies done so far also suggest that children's long-COVID symptoms could be similar to those seen in adults. These findings need to be confirmed by more detailed surveys.

Relatively few such studies are in the works. Last year, the US National Institutes of Health announced that it would set aside US\$1 billion for research into long COVID. One of these is a study to track the recovery of a 'metacohort' of 40,000 adults and children infected with SARS-CoV-2. In the United Kingdom, the National Institute for Health Research, the main funding agency for health science, has funded three studies of long COVID that include younger people. One is the CLoCk study. A second is exploring

how families are affected by the condition. The third is a study of the immune systems of people with long COVID, intended to elucidate the underlying mechanisms of the condition and suggest potential treatments. The study is mostly focused on adults, but includes a small number of children.

The disparity is also noticeable in trials of potential treatments for long COVID. Of several that are under way, none involves teenagers or children. This reflects a general pattern in medical science: adults are studied first and children come later, partly for safety reasons, so that therapies can be tested on adults before they are tested on children.

It is, of course, harder to obtain data for children below the age of 11, and there are legitimate challenges in recruiting children for trials, including obtaining informed consent from parents or guardians. But it's important that young people are not forgotten. Institutions and funding agencies need to think harder and more creatively, otherwise children with conditions such as long COVID will continue to be left behind.

Nature **602**, 183 (2022)

doi: <https://doi.org/10.1038/d41586-022-00334-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00334-w>

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

- EDITORIAL
- 09 February 2022

Africa is bringing vaccine manufacturing home

A major milestone was reached last week when scientists in South Africa reproduced Moderna's COVID-19 vaccine. COVID-19 patents must now be shared.



Scientists at Afrigen Biologics and Vaccines (pictured) have shown that mRNA vaccines can quickly be replicated. Credit: Rodger Bosch/AFP/Getty

It took a pandemic for the world to learn that [African countries import 99% of their vaccines](#). Africa has around ten vaccine manufacturers, but most do not make a vaccine's active ingredients, and instead 'fill and finish' imported products. A lack of manufacturing is one reason that only 11% of the continent's people have been fully vaccinated against COVID-19.

African leaders, meeting for their annual summit in Addis Ababa last week, reiterated a target of vaccinating 70% of their populations this year. But the World Health Organization (WHO) has said that this would need a sixfold increase in weekly vaccinations — from 6 million people to 36 million. So far, only Mauritius and Seychelles have met the 70% target, and COVAX, an initiative to provide vaccines for low- and middle-income countries, is running out of money.

African countries continually find themselves at the back of the vaccine queue, but two developments could begin to change this narrative. Last week, researchers at a company in South Africa said that they have [nearly completed the process of reproducing Moderna's mRNA vaccine against COVID-19](#). Working with the WHO's technology-transfer hub, the researchers at Afrigen Biologics and Vaccines in Cape Town made very small quantities of vaccine, based on Moderna's data, but without the company's involvement. The WHO advised them to copy Moderna's vaccine in part because the company, based in Cambridge, Massachusetts, has said it will not enforce its COVID-19 patents during the pandemic.

Deciding when the pandemic is over, and whether next-generation mRNA vaccines might infringe particular patents, is fraught with difficulty. But a second development would make life easier for smaller manufacturers that can make mRNA vaccines and scale up production. World Trade Organization (WTO) member states might be edging closer to an [agreement to waive intellectual-property rights](#) for COVID-19 vaccines and treatments during the pandemic.

This idea, proposed by India and South Africa, has the backing of more than 100 nations, as well as that of researchers, campaign groups, businesses and media outlets, including *Nature*. The European Union, Switzerland and the United Kingdom are opposed, but WTO director-general Ngozi Okonjo-Iweala is working to resolve the differences. One possible compromise could

be for the waiver to apply only in countries that lack vaccine manufacturing and research.

Countries opposing the waiver argue that it is unlikely to change the course of the pandemic, partly because transferring knowledge and ramping up manufacturing capacity would take years. But the achievements of the researchers in South Africa show that this is not the case. Moreover, the waiver's function is to catalyse manufacturing where this is lacking, create greater self-sufficiency and ensure that vaccines reach vulnerable populations more quickly. Pharmaceutical companies are concerned about losing profits, but researchers who study the economics of the industry argue that it should consider giving something back, considering the scale of public funding for COVID-19 vaccines ([L. Hawksbee et al. BMJ 376, e067367; 2022](#)).

Companies need to accept that the WTO talks seem to be inching towards a consensus. African Union member states, moreover, are implementing a target announced last April for 60% of Africa's routinely used vaccines to be made on the continent within the next two decades. The African Development Bank plans to invest up to US\$3 billion to support the pharmaceutical industry over 10 years, the bank's vice-president for infrastructure, Solomon Quaynor, told *Nature*. This will include funding for transport and logistics infrastructure, and supporting Africa's capacity for medicines regulation. Around \$100 million will be invested annually in pharmaceuticals manufacturing, including vaccines. The aim is to support two to three companies or projects each year. In addition, BioNTech in Mainz, Germany, which produced an mRNA vaccine with New York-based Pfizer, has announced that it will build manufacturing plants in Rwanda and Senegal — welcome recognition that something is wrong when vaccine manufacturing is almost completely absent from an entire continent.

Greater transparency

The large manufacturers have been supplying COVAX with vaccines, and that has been crucial for Africa, with nearly 60% of doses coming from this one source. Last month, it administered its billionth dose, but it is short of funds and urgently needs \$5 billion to distribute doses. One of COVAX's

biggest challenges has been that neither governments nor suppliers have been fully transparent. Companies tend to keep details of vaccine doses and prices confidential. The problem with this is that a company, having agreed to supply COVAX, can still prioritize a higher-paying customer. COVAX is left waiting, and no one will know why. If COVAX is to improve vaccine supply, governments should publish data on prices paid and delivery time-tables. Companies want to keep prices secret so that their competitors don't know what they are charging. But this might be a short-lived benefit, because that information will eventually come out in national audits.

COVAX has been essential to Africa's fight against COVID-19. But, ultimately, countries should not have to rely on aid for vaccination. African countries must be supported in their efforts to own the process, through patent waivers and support for boosting capacity, including that from the WHO's technology-transfer hub. A [special collection of articles in the latest issue of the journal *Nature Human Behaviour*](#) shows clearly that equitable vaccination is within reach — and that it benefits everyone. The efforts of scientists in South Africa and their collaborators around the world should be valued and applauded: they have shown how vaccine production can happen on any continent.

Nature **602**, 184 (2022)

doi: <https://doi.org/10.1038/d41586-022-00335-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00335-9>

- WORLD VIEW
- 08 February 2022

Tracking COVID-19 infections: time for change



We need better numbers if we are to manage the pandemic.

- [Natalie Dean](#) 0

One of the best ways the world has to get a clear view of COVID-19 is going underused. It's time to exploit the power of random sampling.

Last September, the US Centers for Disease Control and Prevention estimated that only one in four SARS-CoV-2 infections in the United States had been reported. Across Africa, the average is closer to one in seven. Why? Many people who are quite ill, or worried about their symptoms, can't get tested. Those with mild or no symptoms often don't seek testing.

And undercounts are getting worse. Reinfections and breakthrough infections are rising, but they are often mild, so people go untested. The onslaught of Omicron cases has far outstripped many countries' testing capacities. Last December, a testing site near me in Atlanta, Georgia, had a wait of three to four hours. In the United States, at-home lateral flow tests are finally becoming more readily available, so fewer people will seek PCR confirmation.

All this undercounting renders many important questions unanswerable. For example, if a surge in cases slows, is transmission down, or is testing maxed out? Waiting to find out means that hospitals can't prepare and policymakers are two to four weeks behind. Who can drive looking only in their rear-view mirror?

Wastewater surveillance is an innovative part of the solution. It shows whether virus levels are increasing or decreasing across a community, and does not depend on people seeking or reporting test results. In my home state of Massachusetts, waste water was one of the earliest reliable indicators that infections were declining last month.

But waste water can't pinpoint who in a community is getting infected and who is getting sick. With Omicron, hospitalizations in children have reached record highs. Yet infections in this age group are frequently missed. It's clear there are more infections, but are those infections more severe? Knowing that is important for risk–benefit calculations around schooling, vaccinations and much more.

Random sampling can answer those sorts of question. As long as participants are selected randomly, they will on average mimic characteristics of the wider population. Roughly speaking, testing fewer than 1,000 people can yield crucial information about 10 million, or even more.

Shining examples of random sampling are the Coronavirus (COVID-19) Infection Survey run throughout the United Kingdom by the Office of National Statistics (ONS), and Imperial College London's REACT-1 study. The ONS initiative aims to obtain swab test results at least fortnightly from around 180,000 people across the United Kingdom, and blood tests monthly

from around 150,000 people. In late January, one in 20 people tested positive for current infection. But age really mattered: one in 10 of the youngest children tested positive, as did one in 15 of the older children. The results signalled an enormous pool of infections, and were quickly made available to guide policy and family decisions.

Forecasting the course of the pandemic demands reliable estimates of current infection levels. Without accurate knowledge of these levels, epidemiologists must make many assumptions (on the likelihood that, for example, infected people will develop symptoms, or be tested). That guesswork informs mathematical models and, consequently, public discussions about the trajectory of the pandemic. Models that overestimate how many infections have been missed overestimate population immunity, and can underestimate the risk of resurgence. Those estimates are used for decisions about everything from opening schools to planning policies and targeting vaccination campaigns. Without random sampling, there's a vicious cycle of guesswork.

The UK data are informative elsewhere, but generalizing too much from one country's data is perilous. In the United States, a few random-sampling surveys have been conducted by health departments and academic partners, for example, in Indiana, Georgia and California. These have bolstered local understanding of disparities across racial and ethnic groups. At a national level, researchers at Emory University in Atlanta (where I also work), carried out a representative household survey ([P. S. Sullivan et al. *Clin. Infect. Dis.* <https://doi.org/hfvm>; 2021](https://doi.org/hfvm)). A new round of antibody and nasal-swab testing is conducted every four to nine months. But a situation that's evolving quickly requires more frequent samples.

Why isn't random-sampling for infection happening more widely? These studies require sustained resources and coordinated effort. The patchwork US public-health system makes collaboration across states challenging. The studies also require a public that's willing and able to participate. Low participation rates in surveys are a major challenge. As an incentive to take part in the ONS survey, the UK government has offered more than £200 million (US\$270 million) of shopping vouchers.

More than two years into the COVID-19 pandemic, it is clear that the virus SARS-CoV-2 will be circulating for a long time to come. Millions of people are being infected daily, and the threat of new variants looms. Investing in random sampling can better prepare governments for the future. A single sampling framework can be used for multiple pathogens, such as influenza and other respiratory viruses. For infectious diseases, failing to see the whole picture will mean poor decisions. Yes, random sampling will cost, but bad information is expensive, too.

Nature **602**, 185 (2022)

doi: <https://doi.org/10.1038/d41586-022-00336-8>

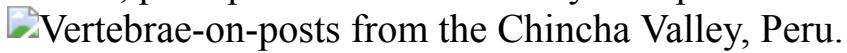
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00336-8>

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

- RESEARCH HIGHLIGHT
- 02 February 2022

‘Spines on posts’ hint at ancient devotion to the dead

Indigenous people in what is now Peru partially re-assembled human skeletons, perhaps after desecration by European tomb robbers.



Nearly 200 reed posts strung with human vertebrae have been documented in the Chincha Valley of western Peru. Credit: Colleen O’Shea

Centuries ago, people living on South America’s Pacific Coast stacked the remains of human spines onto reed sticks — a practice that could have served to piece together the damaged skeletons of their dead¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **602**, 186 (2022)

doi: <https://doi.org/10.1038/d41586-022-00234-z>

References

1. Bongers, J. L., Mejía, J. G., Harper, T. K. & Seidensticker, S. *Antiquity* <https://doi.org/10.15184/aqy.2021.180> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00234-z>

- RESEARCH HIGHLIGHT
- 26 January 2022

Killer whales teach pals a new snack source: fishing lines

Orcas in one area of the southern Indian Ocean now snatch some 180 tonnes of valuable Patagonian toothfish from fishing lines every year.

A killer whale (*Orcinus orca*) diving underwater at Possession Island, part of the Subantarctic Crozet Archipelago.

A killer whale near the Crozet Islands in the southern Indian Ocean. Local whales have learnt to remove fish from commercial fishing lines. Credit: Xavier Desmier/Gamma-Rapho via Getty

Killer whales have invented a behaviour and are teaching it to one another¹, much like how humans once shared bronze smelting or how Internet users today spread the latest TikTok dance challenge.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 186 (2022)

doi: <https://doi.org/10.1038/d41586-022-00233-0>

References

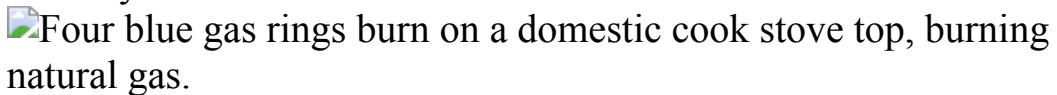
1. Amelot, M. *et al.* *Biol. Lett.* <https://doi.org/10.1098/rsbl.2021.0328> (2022).

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

- RESEARCH HIGHLIGHT
- 04 February 2022

Gas stoves help to cook the climate — even when off

US kitchen appliances that burn natural gas emit much more methane than previously realized.



Gas stoves are good for making tea but bad for the climate — they leak substantial amounts of methane to the atmosphere. Credit: Angel Garcia/Bloomberg via Getty

Gas-burning stoves and ovens in the United States alone could be adding half-a-million-cars' worth of greenhouse gases to the atmosphere every year¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 186 (2022)

doi: <https://doi.org/10.1038/d41586-022-00240-1>

References

1. Lebel, E. D., Finnegan, C. J., Ouyang, Z., & Jackson, R. B. *Environ. Sci. Technol.* <https://doi.org/10.1021/acs.est.1c04707> (2022).

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

- RESEARCH HIGHLIGHT
- 31 January 2022

Microbes that inflict deadly diarrhoea gain new powers

Shigella bacteria, which kill more than 200,000 people every year, are evolving resistance to crucial antibiotics.

 Coloured scanning electron micrograph (SEM) of *Shigella flexneri* bacteria.

Shigella flexneri bacteria (artificially coloured), which are a common cause of diarrhoea in children in low- and middle-income countries, have a large and diverse array of antibiotic-resistance genes. Credit: BSIP VEM/Science Photo Library

Shigella bacteria cause nearly one in six diarrhoea-related deaths worldwide and are the top cause of severe childhood diarrhoea in low- and middle-income countries. These pathogens are becoming highly resistant to key antibiotics, particularly in regions where the drugs are widely used — making them even more dangerous.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

[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.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 186 (2022)

doi: <https://doi.org/10.1038/d41586-022-00235-y>

References

1. Bengtsson, R. J. *et al.* *Nature Microbiol.* <https://doi.org/10.1038/s41564-021-01054-z> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00235-y>.

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

- RESEARCH HIGHLIGHT
- 01 February 2022
- Correction [03 February 2022](#)

Rare metal helps to turn sunlight into fuel, day and night

Molecule that incorporates a light-sensitive ruthenium compound can store solar energy, then release it later to form hydrogen fuel.

A hydrogen fueling pump lit at night at a TrueZero station in Mill Valley, California, U.S.

A fuel pump in California dispenses not petrol, but hydrogen — which could now be produced in the dark, despite depending on solar power, thanks to a newly devised molecule. Credit: David Paul Morris/Bloomberg via Getty

A molecule can harness solar power to generate hydrogen, a clean fuel, on demand — even when it's dark¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 187 (2022)

doi: <https://doi.org/10.1038/d41586-022-00232-1>

Updates & Corrections

- **Correction 03 February 2022:** An earlier version named the wrong journal as the publisher of this research.

References

1. Amthor, S. *et al.* *Nature Chem.* <https://doi.org/10.1038/s41557-021-00850-8> (2022).

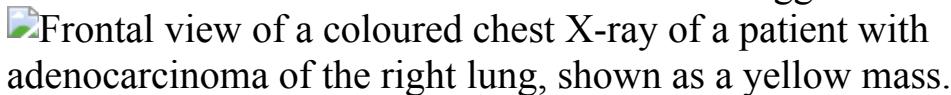
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00232-1>

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

- RESEARCH HIGHLIGHT
- 03 February 2022

Huge cancer database yields genetic clues to metastasis

Scientists sift through samples from more than 25,000 people to identify a host of mutations that could make tumours more aggressive.



A type of tumour called a lung adenocarcinoma (yellow; artificially coloured) is more likely to spread if it contains certain genomic changes.
Credit: Pr Michel Brauner/ISM/Science Photo Library

Scientists have identified a trove of genetic mutations that are linked to the development of metastatic cancer¹. These mutations could act as cancer cells' navigation apps, providing directions that allow malignancies to spread from one organ to another.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 187 (2022)

doi: <https://doi.org/10.1038/d41586-022-00236-x>

References

1. Nguyen, B. *et al.* *Cell* <https://doi.org/10.1016/j.cell.2022.01.003> (2022).

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

- RESEARCH HIGHLIGHT
- 03 February 2022

How to protect unvaccinated children from COVID: vaccinate their parents

Data from more than 150,000 households show that jabs for adults drastically reduce young family members' risk of SARS-CoV-2 infection.



A COVID-19 vaccination centre in Kolkata, India. Vaccinating parents has the added bonus of reducing their children's risk of contracting SARS-CoV-2. Credit: Dipayan Bose/NurPhoto via Getty

Parents vaccinated against COVID-19 shield their unvaccinated children from infection with the Alpha and Delta variants of SARS-CoV-2, according to a study of more than 150,000 households in Israel¹.

More than two-thirds of Israel's residents have received the Pfizer–BioNTech jab. To study the vaccine's full effects, Noam Barda at Ben Gurion University of the Negev in Beersheba, Israel, and his colleagues analysed health-record data collected during two waves of COVID-19, when schools were partially closed.

Between January and March 2021, when Alpha was dominant, children under the age of 16 were around 72% less likely to get infected if they lived with two fully vaccinated parents than if they lived with two unvaccinated parents. Between July and September 2021, when Delta was dominant, children younger than 12 were 58% less likely to become infected if both

parents had received a booster jab than if they had received only two vaccine doses.

The vaccines probably protected kids by reducing the risk of parents getting infected in the first place. If they did get infected, vaccinated parents were also less likely than unvaccinated parents to spread the infection to their unvaccinated children.

Nature **602**, 187 (2022)

doi: <https://doi.org/10.1038/d41586-022-00239-8>

References

1. Hayek, S. *et al.* *Science* <https://doi.org/10.1126/science.abm3087> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00239-8>

News in Focus

- **[Ten billion COVID vaccines, deadly bacteria and high-risk research](#)** [09 February 2022]
News Round-Up • The latest science news, in brief.
- **[Scientists deliberately gave people COVID — here's what they learnt](#)** [02 February 2022]
News • Only half of participants who were exposed to the coronavirus developed infections, most with mild symptoms.
- **[Does the world need an Omicron vaccine? What researchers say](#)** [28 January 2022]
News • Public-health specialists are debating the need for a shot targeting the variant, now causing a record-breaking surge in COVID-19 cases.
- **[Astronomers close in on new way to detect gravitational waves](#)** [27 January 2022]
News • Several teams hope to use pulsars in the Milky Way to detect ripples in space-time made by distant supermassive black holes.
- **[Last-resort cancer therapy holds back disease for more than a decade](#)** [02 February 2022]
News • Two of the first people treated with CAR-T-cell cancer therapies are still in remission 12 years on.
- **[Social-media platforms failing to tackle abuse of scientists](#)** [28 January 2022]
News • A report by activists found that half of debunked online disinformation targeting three prominent scientists remains live and unlabelled.
- **[The race to save the Internet from quantum hackers](#)** [08 February 2022]
News Feature • The quantum computer revolution could break encryption — but more-secure algorithms can safeguard privacy.
- **[The urine revolution: how recycling pee could help to save the world](#)** [09 February 2022]
News Feature • Separating urine from the rest of sewage could mitigate some difficult environmental problems, but there are big obstacles to radically re-engineering one of the most

basic aspects of life.

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

- NEWS ROUND-UP
- 09 February 2022

Ten billion COVID vaccines, deadly bacteria and high-risk research

The latest science news, in brief.



A teenager gets a shot of a COVID vaccine in Chinhoyi, Zimbabwe, in December. Credit: Tafadzwa Ufumeli/Getty

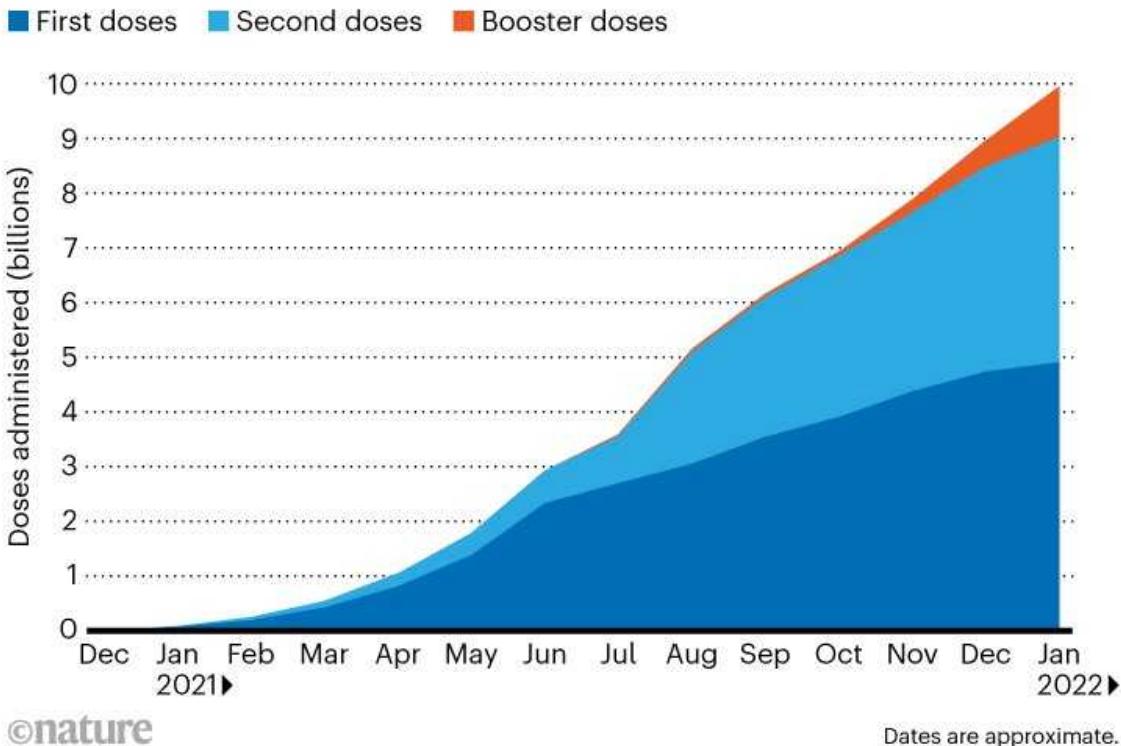
Ten billion COVID vaccinations

In little more than a year, [ten billion doses of COVID-19 vaccines](#) have been administered globally.

Many nations began rolling out vaccines in late 2020 or early 2021, and by late January this year, more than 60% of the global population — 4.8 billion people — had received at least one dose of one of more than 20 COVID-19 vaccines approved for use around the world (see ‘The path to ten billion’).

THE PATH TO TEN BILLION

It took four months after COVID-19 vaccines began to be rolled out to reach one billion vaccinations, but only another nine months to reach ten billion. Almost one billion of these were boosters, raising questions about the inequity facing the large number of people globally who are yet to access even a single dose.



Source: Airfinity

“The world has never seen such rapid scale-up of a new life-saving technology,” says Amanda Glassman, executive vice-president of the Center for Global Development in Washington DC.

But there are still huge inequities in access, with just 5.5% of people in low-income nations having received two doses. By contrast, many of the world’s high- and middle-income nations are now pushing ahead with programmes to deliver third, or even fourth, doses, with these boosters currently making

up around one-third of all COVID-19 vaccine doses administered each day worldwide.

The staggering death toll of drug-resistant bacteria

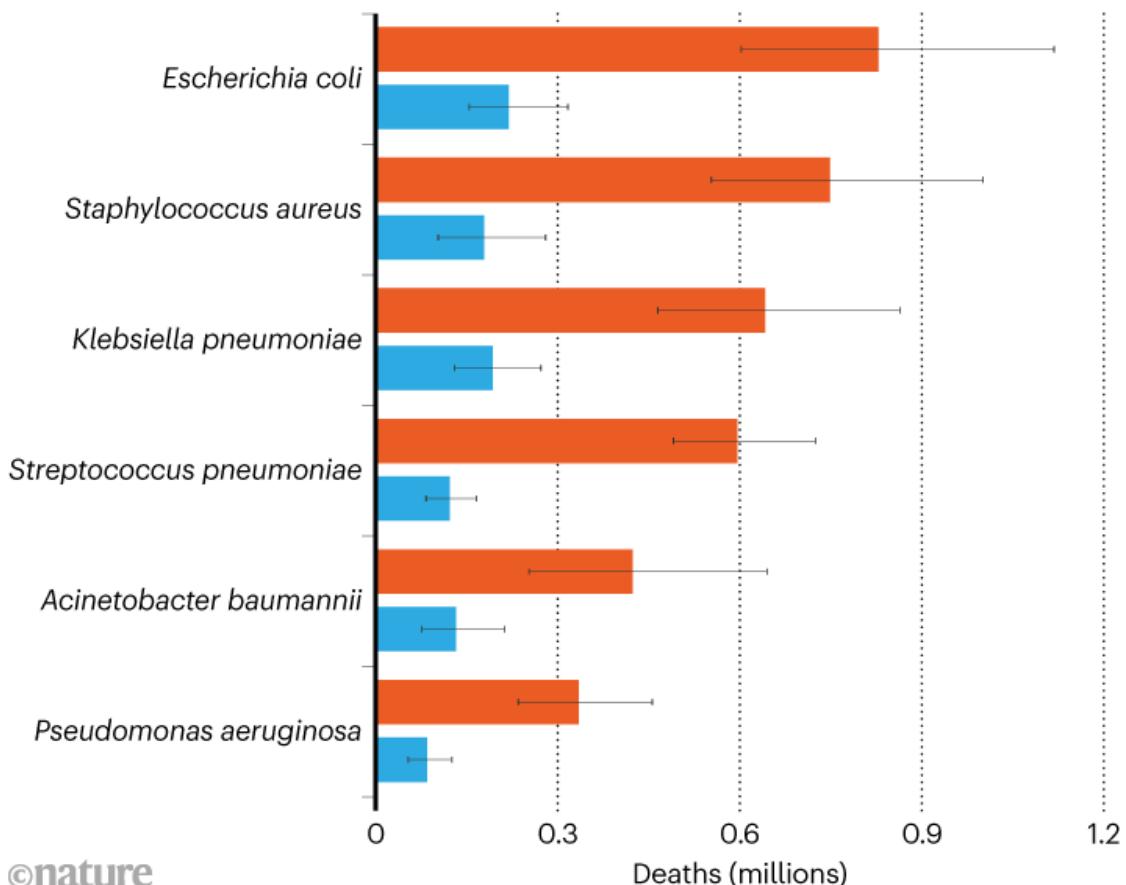
Infections caused by antibiotic-resistant bacteria are [among the leading causes of death globally](#), finds the most comprehensive study yet of antimicrobial resistance (AMR).

The analysis, published on 19 January, estimates that in 2019, 4.95 million people died from illnesses in which bacterial AMR played a part ([C. J. L. Murray et al. Lancet https://doi.org/gn7jdx; 2022](#)). Of those, 1.27 million deaths were the direct result of AMR — meaning that drug-resistant infections killed more people than did HIV/AIDS or malaria.

DEADLY INFECTIONS

These 6 pathogens were responsible for almost 80% of the 1.27 million deaths attributed directly to antimicrobial resistance in 2019.

■ Associated with resistance ■ Attributable to resistance



©nature

Source: Murray, C. L. J. *et al. Lancet* [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0) (2022).

Although there have been many studies on the effects of AMR, few have tried to estimate its global impact. Researchers used data from the 2019 Global Burden of Diseases, Injuries and Risk Factors Study — a survey of 369 diseases and injuries in 204 countries and territories — to estimate the number of people who died from infections that year, along with the pathogens responsible and other factors.

The six deadliest bacterial pathogens were responsible for nearly three-quarters of all deaths attributed to resistance (see ‘Deadly infections’).

Antibiotic-resistant *Escherichia coli* alone killed around 200,000 people in 2019.



Peter Highnam, a computer scientist, has held several posts at US federal research funders. Credit: BEIS

UK's high-risk funder poaches DARPA deputy chief

The UK government has announced the [first chief executive of its new high-risk, high-reward research-funding agency](#): Peter Highnam, the deputy director of the US Defense Advanced Research Projects Agency (DARPA), on which the British organization is modelled. Highnam will start his five-year post in May.

The UK agency, known as the Advanced Research and Invention Agency (ARIA), will have a budget of £800 million (US\$1 billion) over 4 years. It will give Highnam, and whoever chairs the organization, the power to

choose which areas of science to fund. This approach stands in contrast to UK Research and Innovation, the country's main research-funding organization, which disburses its yearly budget mostly through competitive grant schemes.

Highnam has worked at DARPA since 2018, and he has stepped up run the agency temporarily. He has previously worked at the US National Geospatial-Intelligence Agency and the Intelligence Advanced Research Projects Activity, a DARPA spin-off that funds research to help the intelligence community. His appointment follows a flurry of activity in setting up ARIA. Last month, the bill that lays the legislative groundwork to set up the agency passed through UK Parliament.

Nature **602**, 189 (2022)

doi: <https://doi.org/10.1038/d41586-022-00337-7>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00337-7>

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

- NEWS
- 02 February 2022

Scientists deliberately gave people COVID — here's what they learnt

Only half of participants who were exposed to the coronavirus developed infections, most with mild symptoms.

- [Ewen Callaway](#)



Study participants reported symptoms typical of other respiratory infections, such as runny noses and sore throats. Fevers were less common.Credit: Morteza Nikoubazl/NurPhoto via Getty

Healthy, young people who were intentionally exposed to the coronavirus SARS-CoV-2 developed mild symptoms — if any — in a first-of-its-kind COVID-19 human-challenge study. Such trials present a unique opportunity to study viral infections in detail from start to finish, but are controversial because of the risks they pose to participants.

The UK study of 34 individuals, aged 18–30 years, shows that such trials can be done safely, say scientists, and lays the groundwork for more in-depth studies of vaccines, antivirals and immune responses to SARS-CoV-2 infection. The results were posted¹ on 1 February on the preprint server Research Square and have not been peer reviewed.

Nearly half of the participants who received a low dose of virus did not become infected, and some of those who became infected had no symptoms. Participants who did develop COVID-19 reported mild-to-moderate symptoms, including sore throats, runny noses and loss of smell and taste.

“It presents a potentially important advance in how to assess future vaccine and drug efficacy,” says Miles Davenport, an immunologist at the University of New South Wales in Sydney, Australia. “This opens a number of important possibilities to study immunity in a controlled environment.”

However, some researchers question whether the insights yielded by the study so far are important enough to justify the risks to participants, such as the potential for long-term side effects. “In my mind, it’s still not entirely clear whether these studies are ethically justified, and I’m waiting to see what else they’ve found,” says Seema Shah, a bioethicist at Northwestern University in Chicago, Illinois.

Finding the dose

Human-challenge studies have been used for decades to study influenza, malaria and numerous other infectious diseases. Some researchers argued in favour of conducting such trials with SARS-CoV-2 in the early months of the pandemic, as a way to accelerate the development of vaccines. But others saw challenge trials as too dangerous to be acceptable, when so little

was known about the virus and few, if any, effective treatments were available.



The challenge trial was conducted in London. Credit: hVIVO

The trial, led by researchers at Imperial College London and a Dublin-based commercial clinical-research organization called Open Orphan and its London-based subsidiary hVIVO, was announced in October 2020, and the first participants were exposed to the virus in early 2021. Volunteers received £4,565 (US\$6,200) for their participation, which involved at least two weeks of quarantine in a high-level isolation unit at the Royal Free Hospital in London.

The first participants received a very low dose — roughly equivalent to the amount of virus in a single droplet of nasal fluid — of a virus strain that circulated in the United Kingdom in early 2020. Researchers anticipated that a higher dose would be needed to infect a majority of participants, says Andrew Catchpole, chief scientific officer of hVIVO. But the starting dose successfully infected more than half of the participants.

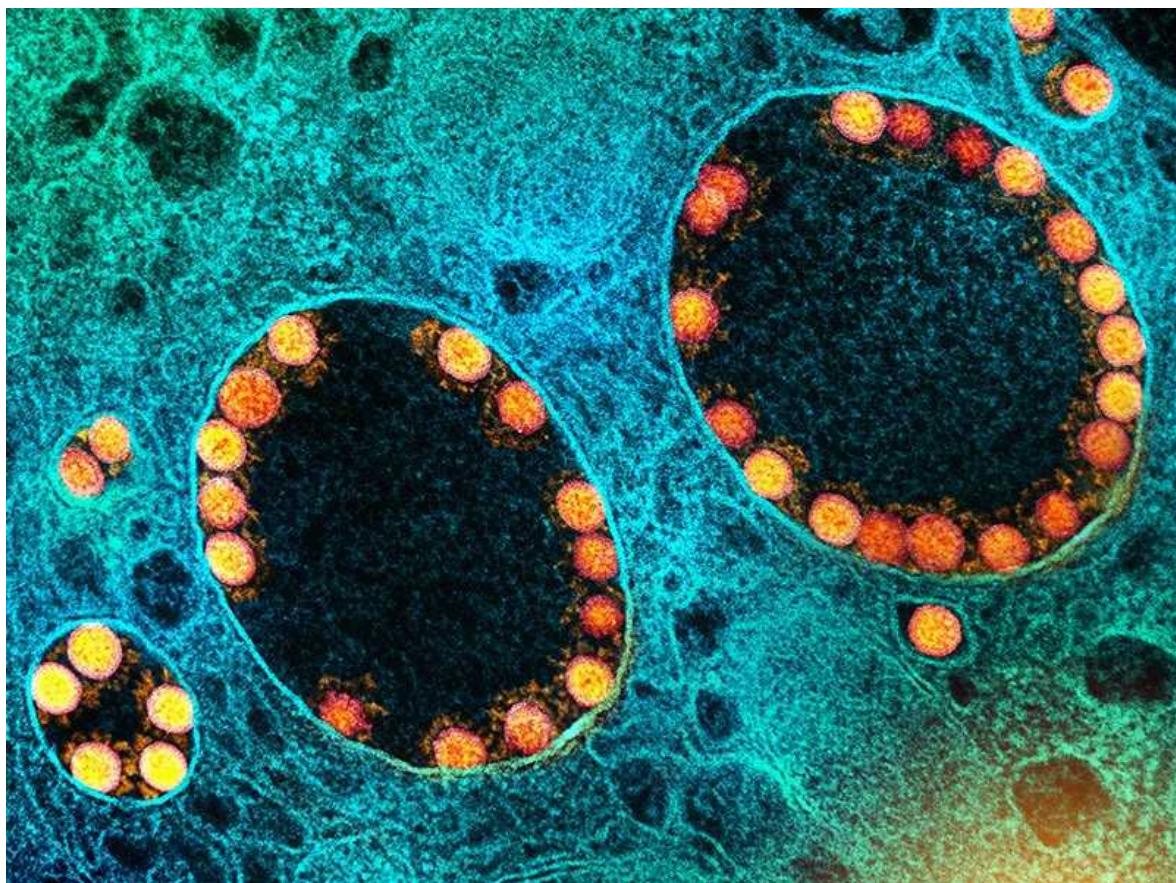
The virus replicated incredibly rapidly in those who became infected. On average, people developed their first symptoms and tested positive, using sensitive PCR tests, less than two days after exposure, on average. That

contrasts with the roughly five-day ‘incubation period’ that real-world epidemiological studies have documented between a probable exposure and symptoms. High viral levels persisted for an average of 9 days, and up to 12 days.

The most common symptoms were typical of other respiratory infections: sore throats, runny noses and sneezing. Fever was less common, and no one developed the persistent cough that had been used as a hallmark of COVID-19, says Catchpole. Around 70% of infected participants lost their senses of smell or taste — another COVID-19 signature — to varying degrees. Such problems persisted for more than six months in five participants and more than nine months in one. Some people developed no symptoms at all, but had as much virus in their upper airways as did participants who exhibited symptoms, and their infections lasted for as long.

Researchers involved in the study want to understand why so many people did not become infected, despite being exposed to SARS-CoV-2. Some uninfected participants had very low levels of virus for short periods of time, suggesting that their immune systems were actively fighting the virus, says Christopher Chiu, a physician-scientist at Imperial College London, who led the study.

Future studies of the challenge-trial participants will attempt to explain why. Previous research has suggested that coronaviruses that cause the common cold might confer protection against COVID-19 in some people. Another possibility is that some people have potent innate immune responses that don’t require a previous encounter with a particular pathogen or a closely related virus. “We’re trying to understand the fundamentals of why people are protected even though they’ve not been exposed to a virus like this before,” Chiu adds.



A micrograph of SARS-CoV-2 virus particles (gold). Credit: NIAID ([CC BY 2.0](#))

His team plans to launch another challenge trial that will expose vaccinated people to the Delta variant of SARS-CoV-2. That study will attempt to identify immune factors that protect people from ‘breakthrough’ infections after vaccination. For the time being, human-challenge trials for SARS-CoV-2 will probably enrol only people at very low risk of severe disease, says Catchpole. But as researchers gain experience running these challenge trials safely, it might be possible to expand them to involve at-risk groups, such as older people, Chiu adds.

Concerns linger

The study looked safe and well-conducted, says Matthew Memoli, an infectious-disease physician and virologist at the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

It should make some people more comfortable with doing more human-challenge trials for SARS-CoV-2, he adds. Such trials could prove useful in the development of vaccines that protect against a broad range of coronaviruses, not just SARS-CoV-2, he adds.

Meagan Deming, a vaccine scientist and virologist at the University of Maryland in Baltimore, says the study confirms insights gained from other COVID-19 studies, such as the swift rise in viral levels. But it has not eliminated her concerns about exposing people to a strain of SARS-CoV-2 that hasn't been weakened. More than two-thirds of participants who became infected had problems with smell or taste that lasted, in some cases, for more than six months, she notes.

"It sounds like this is the most serious risk that materialized. This is the one to keep an eye on," adds Shah. Moreover, she questions whether the insights gleaned from the study so far justify such risks. "This study reads like a promissory note that ultimately, in conjunction with the other research they're doing, there will eventually be substantial scientific and social benefits. But we're not really seeing that yet."

Nature **602**, 191-192 (2022)

doi: <https://doi.org/10.1038/d41586-022-00319-9>

References

1. Killingley, B. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1121993/v1> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00319-9>

- NEWS
- 28 January 2022

Does the world need an Omicron vaccine? What researchers say

Public-health specialists are debating the need for a shot targeting the variant, now causing a record-breaking surge in COVID-19 cases.

- [Emily Waltz](#)



A health-care worker in Lima, Peru, prepares to administer a dose of a messenger-RNA-based COVID-19 vaccine.Credit: Ernesto Benavides/AFP/Getty

After the Omicron variant of SARS-CoV-2 [was first identified in November](#), vaccine makers quickly began developing shots against the highly mutated and transmissible virus. This week, pharmaceutical firm Pfizer and biotechnology company Moderna both announced that they had initiated clinical trials in which they are dosing people with Omicron-based vaccines. But whether rolling out these jabs is necessary, or even practical, is unclear, according to public-health authorities and infectious-disease specialists interviewed by *Nature*.

For those faced with making the complicated decision, debates have begun. Some think it might not be worthwhile because Omicron cases could plummet before manufacturers can finalize the vaccines. Others point out that it's difficult to predict whether the next SARS-CoV-2 variant will be like Omicron, calling into question the utility of an Omicron-specific shot. "We have a lot of confidence in the [current] vaccines, but we must now discuss whether to update the composition," says Kanta Subbarao, who chairs the Technical Advisory Group on COVID-19 Vaccine Composition for the World Health Organization (WHO).

Subbarao's committee formed in September to take a central role in recommending when to update COVID-19 vaccine composition and which variants of SARS-CoV-2 should be targeted. Such a system is already well established for influenza: more than 100 laboratories and 5 WHO collaborating centres across the globe conduct year-round surveillance, testing thousands of virus samples.

The influenza group meets twice per year to recommend the composition of the next season's vaccine, and national regulators and authorities typically follow the WHO's guidance.

So far, no such global structure exists for COVID-19. And until Omicron, it wasn't needed. Vaccine makers based their initial COVID-19 jabs on the SARS-CoV-2 strain first detected in Wuhan, China. But the original virus's biology differs notably from Omicron's, which contains more than 30 mutations in key regions of its genome. Omicron now makes up more than 98% of COVID-19 cases in the United States — one of the hardest-hit nations in the past month — according to the White House COVID-19 Response Team, and is on its way to becoming the dominant variant in

several other countries. The sheer number of cases, and the fact that people are still ending up in hospital, make consideration of an updated vaccine necessary.

Decisions, decisions

How the current vaccines are performing against Omicron is a key factor in the decision. So far, real-world data suggest that a third dose, or booster, of a messenger-RNA-based vaccine, such as those made by Pfizer and Moderna, protects most people infected with Omicron against severe illness, at least in the short term.

The UK Health Security Agency on 14 January reported that a third dose reduces the risk of hospitalization due to Omicron by 92%. The US Centers for Disease Control and Prevention (CDC) on 21 January similarly said that a third shot prevented visits to the emergency room or urgent care with 82% effectiveness, and hospitalization with 90% effectiveness, for people with Omicron¹.

Those numbers are encouraging, but Omicron is spreading rapidly and widely, and many people will still end up in hospital, cautions Emma Thomson, an infectious-disease researcher at the MRC-University of Glasgow Centre for Virus Research, UK. “We shouldn’t underestimate how many people in our communities are vulnerable [to Omicron], either because of age or because of having a weaker immune response for other reasons,” she says.

The protection provided by a booster also wanes quickly, studies show: UK data suggest that effectiveness against hospitalization drops from 92% to 83% just 10 weeks after a third dose.

Decision makers are mulling whether a fourth dose of an original vaccine will be useful, or whether boosting people with a vaccine designed specifically against Omicron makes sense. Essential to those discussions are data on the duration of protection that a third dose provides, says Matthew Hepburn, a senior adviser at the White House Office of Science and Technology Policy (OSTP). “In near-real time, we need to sort out how long

protection lasts, because the fourth-dose conversation is completely shaped by that,” he says. That means watching the vaccines’ effectiveness against Omicron for several more months.

Still, some aren’t waiting that long. In late December, Israel began making a fourth dose of Pfizer’s original vaccine available to its most vulnerable population. And Pfizer and Moderna are pushing forward with Omicron-specific vaccines, now in clinical trials. “I don’t know if or how it will be used,” Pfizer chief executive Albert Bourla said of the Omicron jab to CNBC on 10 January. “But it will be ready.” The company continues to develop the vaccine “as a precaution, but we don’t think it will be needed”, a Pfizer spokesperson told *Nature*. The company expects initial results from the trials in the first half of 2022. Moderna has not provided a timetable.

Chasing variants

If Pfizer meets its ambitious goal — [just months from strain identification to clinical-trial results](#) — it might still be too late to be useful, says Paul Bieniasz, a virologist at Rockefeller University in New York City. Omicron’s dominance as a variant could be waning by then, Subbarao says.

Such a vaccine might work against the variant that dominates after Omicron — especially if the virus continues on that genetic trajectory. But no one knows how the virus will evolve, Bieniasz says. “All of us should be pretty reticent about predicting what is going to be the best matched vaccine months from now,” he says. Just a few months ago, many researchers predicted — logically, but incorrectly — that the next dominant variant to follow Delta would be a virus like Delta. Omicron, however, is completely different.

Semiannual flu vaccine composition decisions are possible because the influenza virus has been around long enough to settle into a manageable pattern — new strains typically derive from previous ones. But that hasn’t been true for SARS-CoV-2. Chasing every variant with a vaccine isn’t tenable, so an alternative approach is needed, say top pandemic experts.

The US National Institute of Allergy and Infectious Diseases (NIAID) and the global Coalition for Epidemic Preparedness Innovations (CEPI), as well as other institutes, have been funding research to develop a pan-coronavirus vaccine that broadly protects against many strains of SARS-CoV-2 and other types of coronavirus. Last year, CEPI announced it would allocate up to US\$200 million in funding for the development of such vaccines, and NIAID has awarded \$42.7 million for such work.

In the meantime, public-health authorities are sorting out how to collaboratively choose an updated vaccine, if one is needed. Such decisions are weighty, considering that there are populations of people in the world who still await access to their first two doses. “There is a lot of fear that if we make a recommendation to update a vaccine, that people might feel that the existing vaccines are no longer useful”, which isn’t true, Subbarao says. That might further delay inoculation for undervaccinated populations. “It’s going to be a difficult message to thread.”

Nature **602**, 192–193 (2022)

doi: <https://doi.org/10.1038/d41586-022-00199-z>

References

1. Thompson, M. G. *et al. Morb. Mortal. Wkly Rep.* **71**, 139–145 (2022).

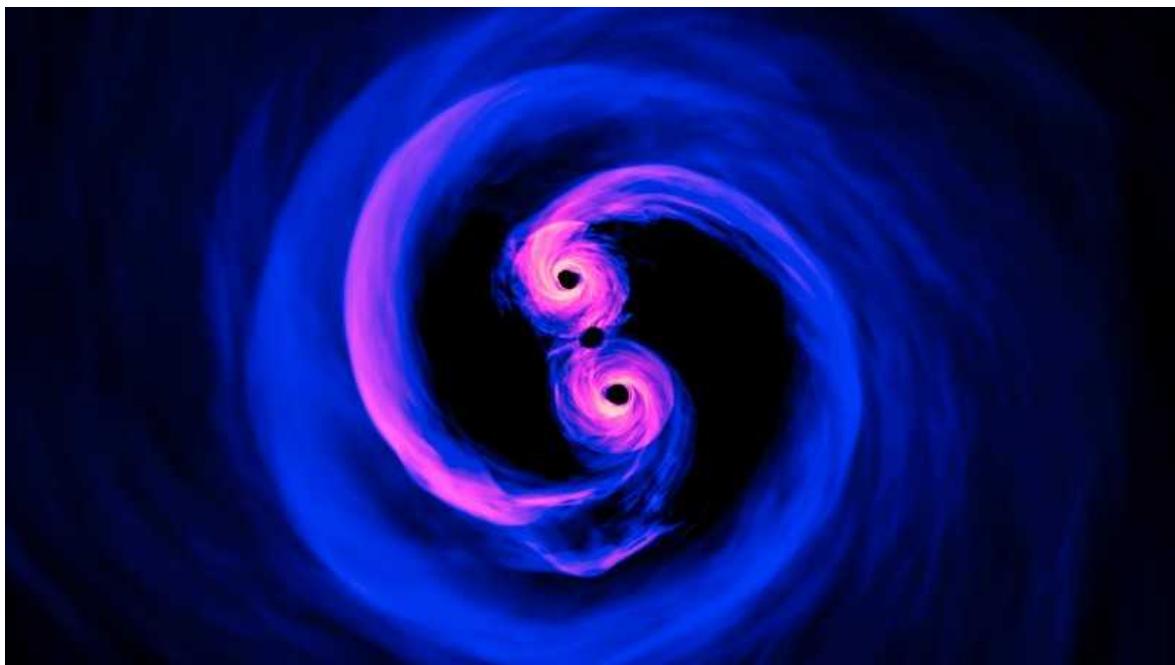
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00199-z>

- NEWS
- 27 January 2022

Astronomers close in on new way to detect gravitational waves

Several teams hope to use pulsars in the Milky Way to detect ripples in space-time made by distant supermassive black holes.

- [Davide Castelvecchi](#)



Supermassive black holes orbiting each other very closely are expected to produce gravitational waves. Credit: NASA's Goddard Space Flight Center/Science Photo Library

Astronomers could be on the verge of detecting gravitational waves from distant supermassive black holes — millions or even billions of times larger than the black holes spotted so far — an international collaboration suggests.

The latest results from several research teams suggest they are closing in on a discovery after two decades of efforts to sense the ripples in space-time through their effects on pulsars, rapidly spinning spent stars that are sprinkled across the Milky Way.

Gravitational-wave hunters are looking for fluctuations in the signals from pulsars that would reveal how Earth bobs in a sea of gravitational waves. Like chaotic ripples in water, these waves could be due to the combined effects of perhaps hundreds of pairs of black holes, each lying at the centre of a distant galaxy.

So far, the International Pulsar Timing Array (IPTA) collaboration has found no conclusive evidence of these gravitational waves. But its latest analysis — using pooled data from collaborations based in North America, Europe and Australia — reveals a form of ‘red noise’ that has the features researchers expected to see. The findings were published on 19 January in *Monthly Notices of the Royal Astronomical Society*¹.

“This is a major milestone,” says Michael Kramer, an astronomer at the Max Planck Institute for Radio Astronomy in Bonn, Germany, who is a leading member of the European team. Although it does not yet constitute a gravitational-wave detection, it is a necessary step towards one, he adds. If the red noise had not been seen at this stage, cosmologists might have had to reconsider their predictions for the prevalence of supermassive black holes and their role in the evolution of the Universe.

Xavier Siemens, a radio astronomer at Oregon State University in Corvallis and a leader of the North American group, agrees that the red noise is not yet a detection. “But it’s reassuring,” he says.

Beyond LIGO

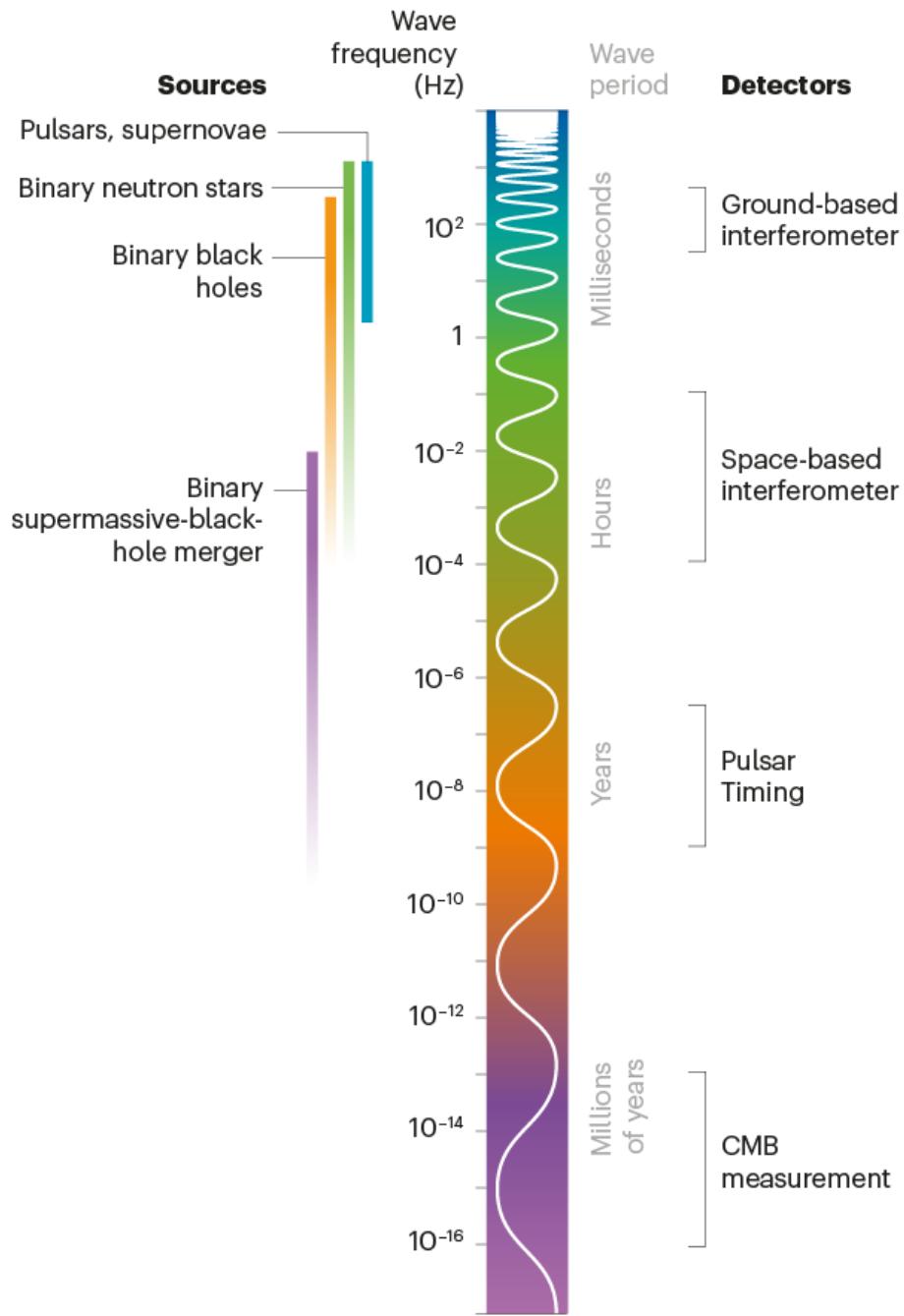
The first direct detection of gravitational waves was achieved in 2015 by the Laser Interferometry Gravitational-Wave Observatory (LIGO) in Louisiana and Washington state. LIGO’s twin antennas measured waves produced in the final moments of the merger of two black holes, each with a mass tens of times that of the Sun. Since then, LIGO and its Italy-based counterpart Virgo

have spotted dozens of similar bursts. Those waves peak in frequency at tens to thousands of cycles per second — similar to the lower frequencies of audible sound — and can be sensed for several seconds or, in some cases, minutes.

The IPTA collaboration's pulsar technique aims to detect longer-lasting gravitational waves that oscillate at much lower frequencies, measured in cycles per year or even per decade (see ‘The gravitational-wave spectrum’). These signals would typically originate from pairs of black holes that orbit each other long-term, rather than from mergers. “This is different from LIGO burst events, where the event occurs very quickly and that particular event will not reoccur,” says radio astronomer George Hobbs at the Australia Telescope National Facility in Epping.

THE GRAVITATIONAL-WAVE SPECTRUM

Much like electromagnetic waves, gravitational waves are emitted by many different objects over a wide range of frequencies. Terrestrial interferometers such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo are sensitive to only a subset of those frequencies, which limits their ability to ‘see’ certain cosmic phenomena. They won’t detect collisions of supermassive black holes found in the hearts of galaxies, for example. But space-based interferometers and other approaches for picking up gravitational waves could extend physicists’ reach.



Astrophysicists think that most large galaxies have a supermassive black hole at their centre. When two galaxies merge, their central black holes eventually sink to the centre of the newly formed galaxy and begin to orbit each other. If they get close enough, they will emit intense gravitational waves.

The pulsar technique looks for these gravitational waves as they sweep through our Galaxy, stretching and compressing the space that separates the Solar System from spinning neutron stars called pulsars (see ‘Pulsars as detectors’). Observatories such as LIGO, by contrast, detect gravitational waves as they sweep Earth.

The approach has unique challenges. Whereas LIGO directly measures minute changes in the distance between two mirrors several kilometres apart, changes in the distance between Earth and a pulsar cannot be measured directly, in part because thousands of gravitational-wave crests and troughs are propagating between them. Earth and the pulsar “are not riding the same crest or trough”, explains Maura McLaughlin, an astronomer at West Virginia University in Morgantown who is a leading member of the North American pulsar collaboration. “To estimate the delay, we have to care about the gravitational waves’ effect on the pulsar and on Earth. The stuff in between cancels out,” says McLaughlin.

Such changes should be revealed because, when local space is stretched, the periodic signals from a pulsar will take tens of nanoseconds more or less to reach Earth than they would have otherwise.

Noisy signals

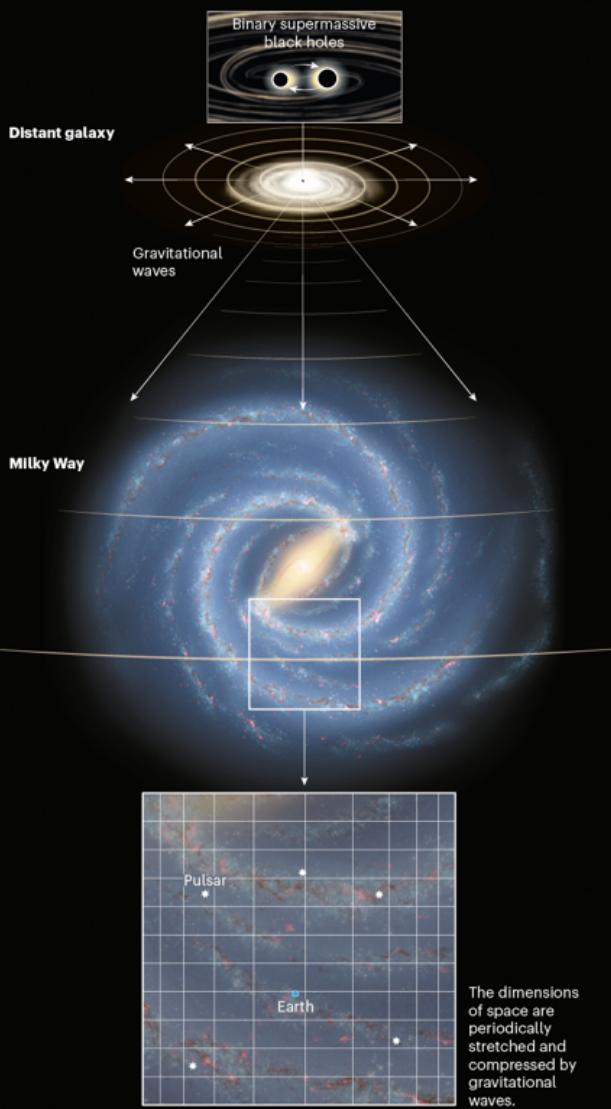
Measuring these delays requires decades of painstaking data gathering, followed by weeks of number crunching on a supercomputer. And it relies on the bizarre physics of the neutron stars known as pulsars.

Many neutron stars — collapsed cores of stars that pack a mass greater than that of the Sun into a sphere just 20 kilometres or so across — spew radiation from their magnetic poles. As a neutron star spins, the beam of radiation circles around like the rotating light of a lighthouse. Some of these

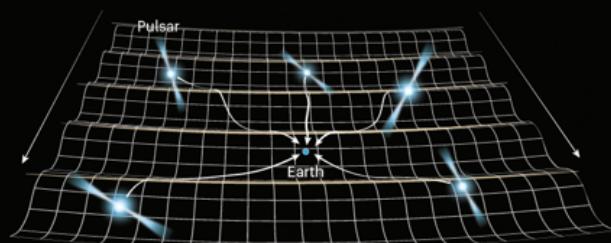
beams happen to cross Earth's path through space, and are detected as radiation pulsating at regular intervals. In the late 1970s, some astronomers pointed out that because they appear at highly regular intervals, some of these beacons could serve as detectors for gravitational waves.

PULSARS AS DETECTORS

The main goal of pulsar timing arrays is to detect gravitational waves produced in distant galaxies by pairs of supermassive black holes that orbit each other closely. The gravitational waves travel for billions of years to reach the Milky Way.



As they stretch and pull space-time at a rate of less than once per year, the waves also affect the propagation of radio signals produced by pulsars. These are neutron stars that act as a magnetic beacon rotating up to 1,000 times per second.



Astronomers have been monitoring dozens of pulsars for decades, looking for tiny frequency changes in their radio signals. Patterns in these changes could reveal the passage of gravitational waves, which shift Earth's position slightly — by tens of metres — with respect to the pulsars.

Nik Spencer/*Nature*; Milky way: NASA/JPL-Caltech/R. Hurt (SSC/Caltech)

But pulsar signals are noisy, and can be slowed or scattered by clouds of interstellar electrons. To overcome this issue, astronomers must compare the signals from as many of these beacons as possible, forming a ‘pulsar timing array’.

And the baseline position of the Solar System’s centre of mass — which is affected by the motions of the planets — must be calculated to a precision of less than 100 metres.

In the past decade, those estimates have improved greatly thanks to measurements of Jupiter and Saturn’s positions made by NASA’s Juno and Cassini missions. The revisions have reassured some astronomers: earlier, less-precise measurements, together with some overly conservative assumptions, had made some worry that the expected gravitational-wave background wasn’t there.

But with each passing year, researchers have accumulated more data and refined their techniques. In 2020 and 2021, each of the three collaborations began to see a telltale sign of the gravitational-wave background^{2,3,4}. Whereas ordinary, ‘white’ noise includes random fluctuations at all frequencies, red noise is louder at lower frequencies. Such a feature is expected when signals of long wavelength — with periods comparable to the 20-odd years of data that have now been accumulated — are beginning to emerge. The IPTA’s latest joint analysis — made by pooling the regional collaborations’ data on 65 pulsars to improve their sensitivity to gravitational waves — has detected the red noise, too, even though it did not use the most recent data sets that the three groups analysed separately in 2020 and 2021.

The finding doesn’t necessarily indicate the presence of gravitational waves. “Red noise can also be produced by other things,” warns Kramer, such as a previously unsuspected pattern in the way the spinning pulsars gradually slow down.

To claim a discovery, “a crucial component is missing”, says radio astronomer Andrea Possenti, a leading member of the European group who

is at the Cagliari Astronomical Observatory in Italy. “These long-term signals must be correlated from one pulsar to another.”

Hobbs agrees. “I personally would like a lot more checks to be done before I’m going to break open the champagne bottle,” he says.

If and when the gravitational-wave background is discovered, “the scientific reward will be immense”, says Monica Colpi, an astrophysicist at the University of Milan–Bicocca in Italy. From the signals, researchers could ultimately get information about how the black holes interacted with dark matter, stars and gas clouds in their galaxies, she says.



The radio telescope at the Arecibo Observatory in Puerto Rico, which collapsed in December 2020, was part of the International Pulsar Timing Array. Credit: Ricardo Arduengo/AFP/Getty

The worldwide effort to hunt for the waves took a hit in December 2020, when the venerable 300-metre Arecibo Observatory — which played an important part in measuring pulsars — [collapsed](#). Since then, the North American team has rerouted some of the work to its other major facility, the

100-metre Green Bank Telescope in West Virginia. “We have dropped a handful of our weaker pulsars and have gaps in our data set of a few months, but all in all we are weathering the loss of [Arecibo] as well as we can,” Siemens says.

Future efforts will benefit from pulsar-timing data being collected at major radio-astronomy observatories in India and South Africa. Eventually, China’s [Five-hundred-meter Aperture Spherical Radio Telescope](#) is expected to join, too.

And the researchers say that the next IPTA paper, expected this year or next, could use the data that already exist to confirm a discovery of the gravitational-wave background produced by supermassive black holes. “Now the time is ripe to bring it all together and make a detection,” says Kramer.

Nature **602**, 194–195 (2022)

doi: <https://doi.org/10.1038/d41586-022-00170-y>

References

1. Antoniadis, J. *et al.* *Mon. Not. R. Astron. Soc.* **510**, 4873–4887 (2022).
2. Arzoumanian, Z. *et al.* *Astrophys. J. Lett.* **905**, L34 (2020).
3. Goncharov, B. *et al.* *Astrophys. J. Lett.* **917**, L19 (2021).
4. Chen, S. *et al.* *Mon. Not. R. Astron. Soc.* **508**, 4970–4993 (2021).

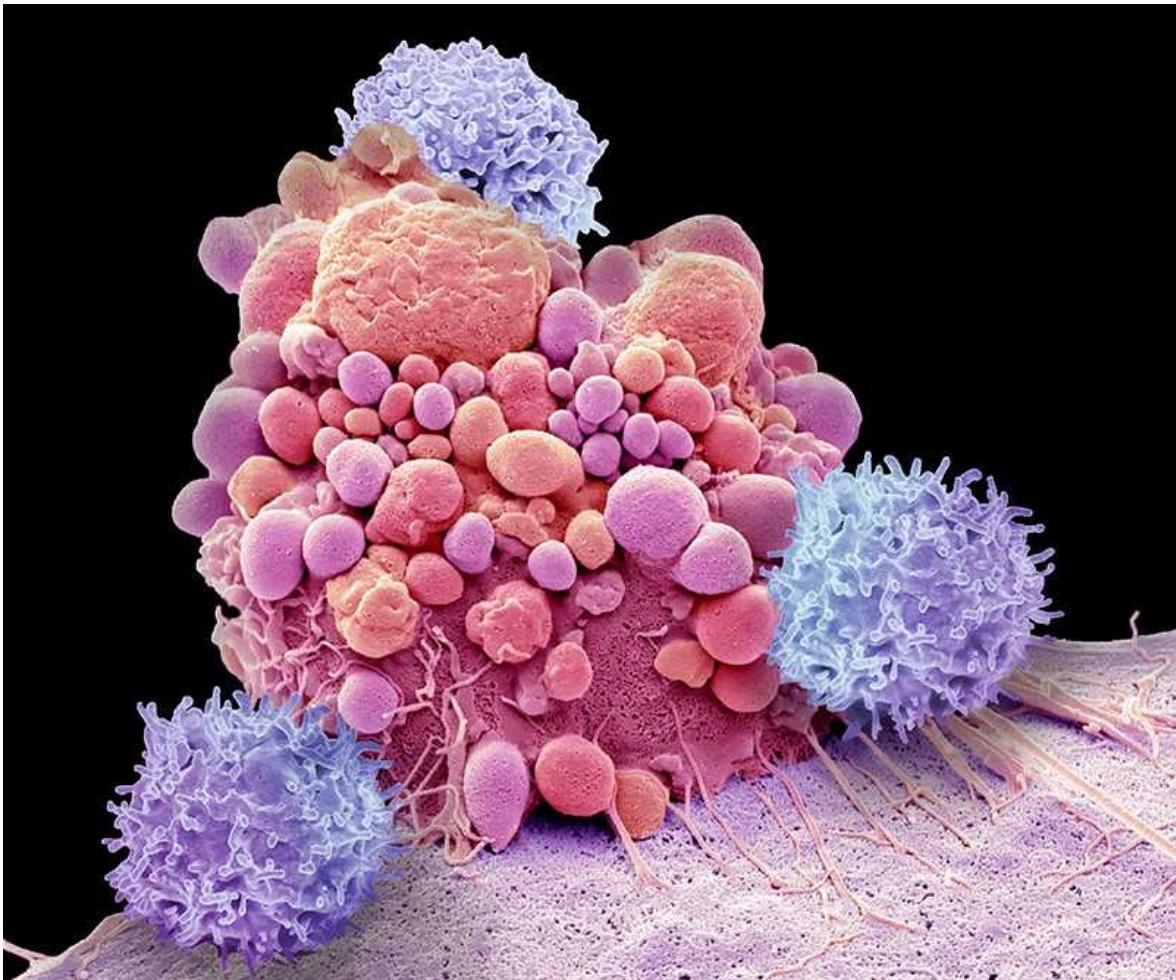
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00170-y>.

- NEWS
- 02 February 2022

Last-resort cancer therapy holds back disease for more than a decade

Two of the first people treated with CAR-T-cell cancer therapies are still in remission 12 years on.

- [Heidi Ledford](#)



CAR-T-cell therapy modifies T cells — shown here attacking a brain cancer cell — to recognize specific proteins.Credit: Steve Gschmeissner/Science Photo Library

A few weeks after receiving an experimental cancer therapy that turns immune cells into tumour-killing hunters, Doug Olson's doctor sat him down to give him news of his progress. "He said, 'Doug, we cannot find a single cancer cell in your body,'" Olson recalls. "I was pretty convinced that I was done with cancer."

Olson's doctors, however, weren't so sure. The year was 2010, and Olson was one of the first people with chronic lymphocytic leukaemia to receive the treatment, called CAR-T-cell therapy. When his doctors — including Carl June and David Porter at the University of Pennsylvania in Philadelphia — wrote the protocol for the clinical trial that Olson was involved in, they

hoped that the genetically engineered cells might survive for a month in his body. They knew that cancer research could be heartbreaking; they didn't dare to expect a cure.

But more than ten years later, the immune cells continue to patrol Olson's blood and he remains in remission. June is finally ready to admit what Olson suspected all along. "We can now conclude that CAR T cells can actually cure patients with leukaemia," June told reporters at a press briefing describing results that were published in *Nature* on 2 February¹.

Tumour destroyers

CAR-T-cell therapies involve removing immune cells called T cells from a person with cancer, and genetically altering them so that they produce proteins — called chimeric antigen receptors, or CARs — that recognize cancer cells. The cells are then reinfused into the person, in the hope that they will seek out and destroy tumours.

In the years since Olson's treatment, five CAR-T-cell therapies have been approved by the US Food and Drug Administration, to treat leukaemias, lymphomas and myelomas. June estimates that tens of thousands of people have received CAR-T cell treatment.

But the therapy is expensive, risky and technically demanding. It remains a last resort, to be used when all other treatments have failed. Despite the treatment's success for Olson, not everyone experiences durable remission of their cancer. In the beginning, only about 25–35% of CAR-T-cell recipients with chronic lymphocytic leukaemia experienced a complete remission of their cancer, says Porter. With refinement, that percentage has increased over the years, he says, but some of these initial successes still lead to relapse. Tracking the treatment long-term could reveal clues as to what factors are important for lasting CAR-T-cell success.

For more than ten years, Porter and his colleagues analysed the CAR T cells in Olson and one other person treated in 2010, tracing the cells' evolution and looking for any signs of safety concerns.

They found that the CAR T cells persisted, but the characteristics of the population shifted over time. Soon after infusion, a prominent population of T cells called CD8⁺ cells emerged. These are sometimes called killer T cells, and can identify and destroy cells that display unusual proteins, such as [cancer cells](#) or [cells that are infected with a virus](#).

But over the years, a different type of CAR T cell became dominant. CD4⁺ T cells can take on a variety of functions in the immune system, but the researchers showed that both study participants had CD4⁺ cells with characteristics suggesting that they would be capable of killing leukaemia cells.

Tremendous impact

Olson and the other participant now have no signs of leukaemia. It's unclear whether the CAR T cells killed all the leukaemia cells soon after they were introduced, or if the cells that continue to patrol are able to destroy any leukaemia cells before they reach detectable levels.

"The potential impact of CAR T is tremendous," says Nirali Shah, a paediatric haematologist at the US National Cancer Institute in Bethesda, Maryland. This study "gives you a proof of concept about the safety of having long-term persistence and integration of the T cells into your body".

It remains to be seen, she adds, how well the findings from these two individuals with chronic lymphocytic leukaemia will translate to other diseases. Efforts are under way to use CAR-T-cell approaches to treat solid tumours, such as prostate tumours and the devastating brain cancer glioblastoma. In January, researchers reported success in using the cells to destroy scar tissue in the heart — an approach that could one day be used to treat cardiac fibrosis².

In the years after his treatment, Olson returned to his career in medical diagnostics. He committed to staying healthy, and his son talked him into running half marathons. "If my cancer was gone, I certainly didn't want to die of a heart attack," he says. Eventually, he decided to go public with the story of his recovery, and serve as a mentor for other people with cancer.

He tries to give them hope, he says: “If there isn’t a cure for their cancer today, there’s a reasonable chance that around the corner, there’s going to be one.”

Nature **602**, 196 (2022)

doi: <https://doi.org/10.1038/d41586-022-00241-0>

References

1. Melenhorst, J. J. *et al.* *Nature* <https://doi.org/10.1038/s41586-021-04390-6> (2022).
2. Rurik, J. G. *et al.* *Science* **375**, 91–96 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00241-0>

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

- NEWS
- 28 January 2022

Social-media platforms failing to tackle abuse of scientists

A report by activists found that half of debunked online disinformation targeting three prominent scientists remains live and unlabelled.

- [Brian Owens](#) ⁰



Anthony Fauci has been targeted with pandemic-related disinformation online. Credit: J. Scott Applewhite/Getty

Social-media sites such as Facebook and Twitter are not doing enough to tackle online abuse and disinformation targeted at scientists, suggests a study by international campaign group Avaaz.

The analysis, [published on 19 January](#), looked at disinformation posted about three high-profile scientists. It found that although all of the posts had been debunked by fact-checkers, online platforms had taken no action to address half of them.

“Two years into the pandemic, even though they have made important policy changes, the platforms, and Facebook in particular, are still failing to take significant action,” says Luca Nicotra, a campaign director for Avaaz who is based in Madrid.

Scientists under attack

Online threats aimed at scientists have become a major problem during the COVID-19 pandemic. [A survey by Nature last year](#) found that many scientists who had spoken publicly about the disease had experienced attacks on their credibility or reputation, or had been threatened with violence. Some 15% had received death threats.

Nicotra and his colleagues looked at pandemic-related disinformation targeting three prominent scientists: Anthony Fauci, head of the US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland; German virologist Christian Drosten; and Belgian virologist Marc Van Ranst. They checked posts across five social-media sites — Facebook, YouTube, Twitter, Instagram and Telegram.

Between January and June 2021, the authors identified 85 posts across the platforms that contained disinformation targeting the scientists and their institutions, and that had been debunked by several fact-checking organizations. By late July 2021, when the study concluded, 49% of the posts were still live and had not been removed or labelled with a warning about the fact-checkers’ findings. The posts had collectively racked up nearly 1.9 million interactions.

The failure to label debunked disinformation is a problem, says Nicotra, because unlabelled posts get much more engagement than ones that are labelled. Labelling is a “very effective strategy” for fighting disinformation, Nicotra says. “Especially if users who have previously interacted with the content are also informed.”

Much of the Avaaz report focuses on Facebook because the platform’s size allows for better statistical analysis, but also because the other sites generally don’t provide access to the necessary data and tools.

“We know enough to say the same problem exists on the others, and it might even be worse,” says Nicotra. “But the lack of transparency makes our job more difficult.”

Problematic posts

A spokesperson for Meta, the parent company of Facebook and Instagram, which is based in Menlo Park, California, says that the company has strict rules on misinformation about COVID-19 and vaccines, and does not allow death threats against anyone on the platforms. It has “removed more than 24 million pieces of content for violating those policies since the pandemic began, including content mentioned in this report”, the spokesperson says. “We’ve also added warning labels to more than 195 million pieces of additional COVID-19 content which don’t violate our policies but are still problematic. We will continue to take action against any of the content that breaks our rules.”

But Nicotra says that the platforms are still missing large numbers of problematic posts, especially outside the United States and Europe, and in languages other than English. In 2020, Facebook devoted just 13% of its budget for developing misinformation-detection algorithms to regions outside the United States, according to documents released by whistleblower Frances Haugen, a former product manager for the company.

Another problem is that the algorithms that govern social media are designed to keep people engaged, and so tend to highlight content that is controversial or emotionally charged, says Nicotra. He says that new

regulations, such as the European Union’s Digital Services Act — which requires companies to assess and act to reduce the risk of harm to society from their products — could force changes to the algorithms.

No silver bullet

“These are underlying problems with social-media platforms that we now see crop up with COVID, and with other crises they will potentially emerge again,” says Heidi Tworek, a historian who studies health communications at the University of British Columbia in Vancouver, Canada.

Although tweaks to algorithms and better enforcement of the companies’ own terms of service will help, Tworek says, there is no silver bullet that will solve the problems of online harassment and misinformation.

Some organizations have started working on ways to support scientists facing online harassment. In December 2021, the Australian Science Media Centre in Adelaide held a webinar that provided practical advice to scientists on how to protect themselves, including how to control privacy settings, and where and how to report abuse. The webinar also highlighted the need for institutions to provide support. “It’s an area that’s often been ignored, but they do have a responsibility of care to their employees,” says Lyndal Byford, the centre’s director of news and partnerships. The UK Science Media Centre (SMC) is planning to run a similar event on 24 February.

Fiona Fox, chief executive of the SMC in London, hopes efforts such as this will help researchers to feel safer talking about their work in public. “We can’t let this stop scientists from engaging with the media,” she says. “The public interest lies in good scientific communication.”

Nature **602**, 197 (2022)

doi: <https://doi.org/10.1038/d41586-022-00207-2>

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

- NEWS FEATURE
- 08 February 2022

The race to save the Internet from quantum hackers

The quantum computer revolution could break encryption — but more-secure algorithms can safeguard privacy.

- [Davide Castelvecchi](#)

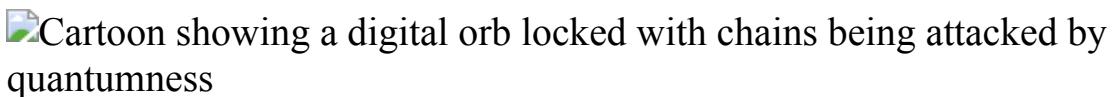


Illustration by Dalbert B. Vilarino

In cybersecurity circles, they call it Q-day: the day when quantum computers will break the Internet.

Almost everything we do online is made possible by the quiet, relentless hum of cryptographic algorithms. These are the systems that scramble data to protect our privacy, establish our identity and secure our payments. And they work well: even with the best supercomputers available today, breaking the codes that the online world currently runs on would be an almost hopeless task.

But machines that will exploit the quirks of quantum physics threaten that entire deal. If they reach their full scale, quantum computers would crack current encryption algorithms exponentially faster than even the best non-quantum machines can. “A real quantum computer would be extremely dangerous,” says Eric Rescorla, chief technology officer of the Firefox browser team at Mozilla in San Francisco, California.

As in a cheesy time-travel trope, the machines that don't yet exist endanger not only our future communications, but also our current and past ones. Data thieves who eavesdrop on Internet traffic could already be accumulating encrypted data, which they could unlock once quantum computers become available, potentially viewing everything from our medical histories to our old banking records. "Let's say that a quantum computer is deployed in 2024," says Rescorla. "Everything you've done on the Internet before 2024 will be open for discussion."

Even the most bullish proponents of quantum computing say we'll have to wait a while until the machines are powerful enough to crack encryption keys, and many doubt it will happen this decade — if at all.

But the risk is real enough that the Internet is being readied for a makeover, to limit the damage if Q-day happens. That means switching to stronger cryptographic systems, or cryptosystems. Fortunately, decades of research in theoretical computer science has turned up plenty of candidates. These post-quantum algorithms seem impervious to attack: even using mathematical approaches that take quantum computing into account, programmers have not yet found ways to defeat them in a reasonable time.

Which of these algorithms will become standard could depend in large part on a decision soon to be announced by the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland.

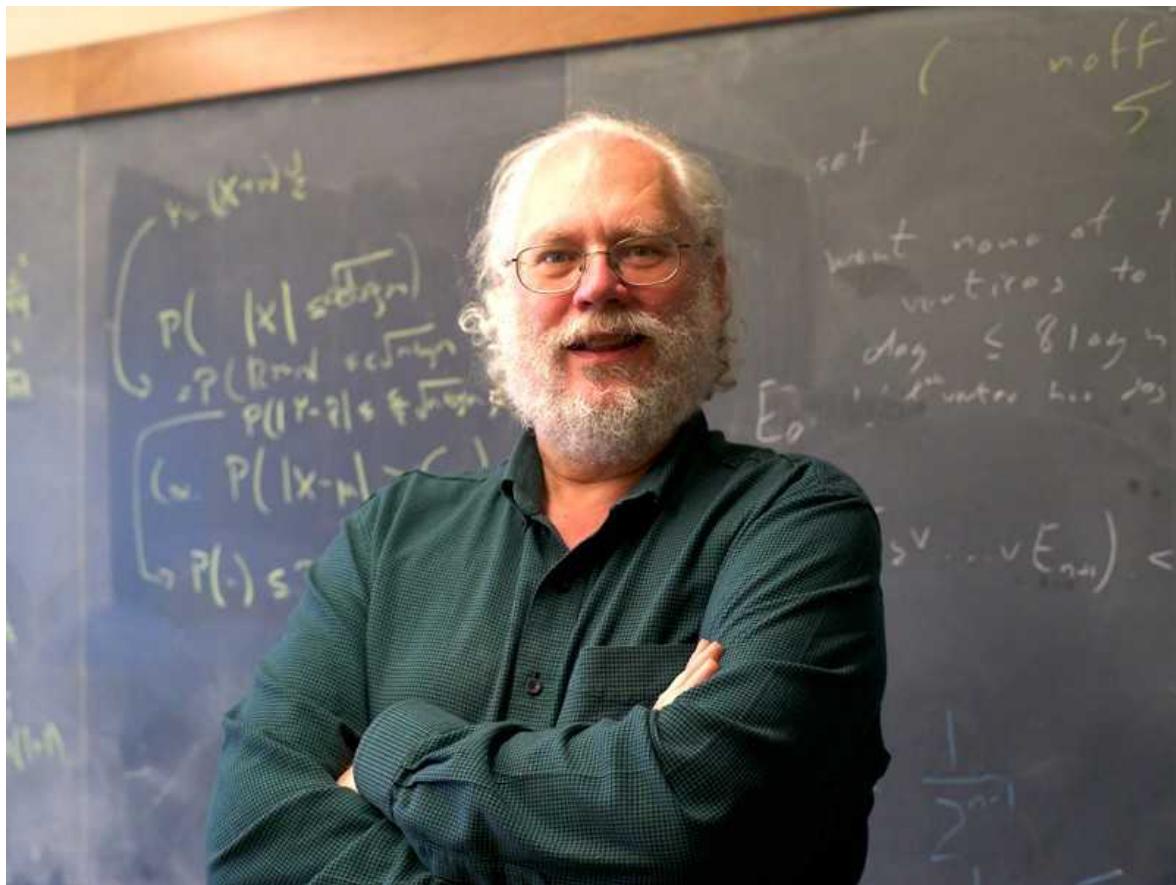
In 2015, the US National Security Agency (NSA) [announced that it considered current cryptosystems vulnerable](#), and advised US businesses and the government to replace them. The following year, NIST [invited computer scientists globally to submit candidate post-quantum algorithms](#) to a process in which the agency would test their quality, with the help of the entire crypto community. It has since winnowed down its list from 65 to 15. In the next couple of months, it will select a few winners, and then publish official versions of those algorithms. Similar organizations in other countries, from France to China, will make their own announcements.

But that will be only the beginning of a long process of updating the world's cryptosystems — a change that will affect every aspect of our lives online, although the hope is that it will be invisible to the average Internet user.

Experience shows that it could be a bumpy road: early tests by firms such as Google haven't all run smoothly.

"I think it's something we know how to do; it's just not clear that we'll do it in time," Peter Shor, a mathematician at the Massachusetts Institute of Technology in Cambridge whose work showed the vulnerabilities of present-day encryption, [told *Nature* in 2020](#).

Even if Q-day never happens, the possibility of code-breaking quantum machines has already changed computer science — and, in particular, the ancient art of cryptography. "Most people I know think in terms of quantum-resistant crypto," says computer scientist Shafi Goldwasser, director of the Simons Institute for the Theory of Computing at the University of California, Berkeley.



Peter Shor showed that quantum algorithms could defeat cryptographic systems. Credit: BBVA Foundation

Birth of public-key cryptography

Armies and spies have always been able to send messages securely even when a channel — be it a messenger pigeon or a radio link — is susceptible to eavesdropping, as long as their messages were encrypted. However, until the 1970s, this required the two parties to agree on a shared secret cipher in advance.

Then, in 1976, three US computer scientists, Whitfield Diffie, Martin Hellman and Ralph Merkle, came up with the revolutionary concept of public-key cryptography, which allows two people to exchange information securely even if they had no previous agreement. The idea rests on a mathematical trick that uses two numbers: one, the public key, is used to encrypt a message, and it is different from the second, the private key, used to decrypt it. Someone who wants to receive confidential messages can announce their public key to the world, say, by printing it in a newspaper. Anyone can use the public key to scramble their message and share it openly. Only the receiver knows the private key, enabling them to unscramble the information and read it.

In practice, public keys are not typically used to encrypt the data, but to securely share a conventional, symmetric key — one that both parties can use to send confidential data in either direction. (Symmetric-key systems can also be weakened by existing quantum algorithms, but not in a catastrophic way.)

For the first two decades of the Internet age from the mid-1990s, the most commonly used public-key-exchange algorithm was RSA, named after its inventors, Ron Rivest, Adi Shamir and Leonard Adleman.

RSA is based on prime numbers — whole numbers such as 17 or 53 that are not evenly divisible by any numbers except themselves and 1. The public key is the product of at least two prime numbers. Only one party knows the factors, which constitute the private key. Privacy is protected by the fact that, although multiplying two large numbers is straightforward, finding the unknown prime factors of a very large number is extremely hard.

More recently, the Internet has been transitioning away from RSA, which is vulnerable even to classical — as opposed to quantum — attacks. In 2018, the Internet Engineering Task Force (IETF), a consensus-based virtual organization that steers the adoption of security standards on a global scale, endorsed another public-key system to replace it. That system is called elliptic-curve cryptography, because its mathematics grew out of a branch of nineteenth-century geometry that studies objects called elliptic curves.

Elliptic-curve cryptography is based on calculating the n th power of an integer (which is associated with a point on the curve). Only one party knows the number n , which is the private key. Calculating the exponential of a number is easy, but given the result, it is extremely hard to find what n was. This technique is faster and more secure than RSA.

All sorts of devices, from mobile phones to cars, use public-key encryption to connect to the Internet. The technology has also spread beyond cyberspace: for example, the radio-frequency chips in everything from credit cards to security passes typically use elliptic-curve algorithms.

Breaking RSA

Just as the number of Internet users worldwide — and the use of public-key cryptosystems such as RSA — was beginning to grow exponentially, Shor, then at AT&T Bell Laboratories in Murray Hill, New Jersey, laid the groundwork for those algorithms' demise. He showed in 1994 how a quantum computer should be able to factor large numbers into primes exponentially faster than a classical computer can ([P. W. Shor Proc. 35th Annu. Symp. Found. Comput. Sci. 124–134; 1994](#)). One of the steps in Shor's quantum algorithm can efficiently break an elliptic-curve key, too.

Shor's was not the first quantum algorithm, but it was the first to show that quantum computers could tackle practical problems. At the time, it was largely a theoretical exercise, because quantum computers were still dreams for physicists. But later that decade, researchers at IBM performed the first proofs of principle of quantum calculations, by manipulating molecules in a nuclear magnetic resonance machine. By 2001, they had [demonstrated that they could run Shor's algorithm](#) — but only to calculate that the prime

factors of 15 are 3 and 5. Quantum-computing technology has made enormous progress since then, but running Shor's algorithm on a large integer is still a long way off.

Still, after Shor's breakthrough, the crypto-research world began to pay attention to the possibility of a Q-day. Researchers had already been studying alternative public-key algorithms, and the news attracted lots of talent to the field, says Goldwasser.

Lattice-based systems

The majority of the algorithms that made it to NIST's final roster rely, directly or indirectly, on a branch of cryptography that was developed in the 1990s from the mathematics of lattices. It uses sets of points located at the crossings of a lattice of straight lines that extend throughout space. These points can be added to each other using the algebra of vectors; some can be broken down into sums of smaller vectors. If the lattice has many dimensions — say, 500 — it is very time-consuming to calculate the smallest such vectors. This is similar to the situation with prime numbers: the person who knows the short vectors can use them as a private key, but solving the problem is extremely hard for everyone else.

Since the 1990s, researchers have developed a plethora of public-key encryption algorithms that either use lattices directly, or are somehow related to them. One of the earliest types, developed in 1996, is called NTRU. Its keys consist of polynomials with integer coefficients, but it is considered secure because of its theoretical similarity to lattice problems. To show that a cryptosystem is trustworthy, researchers often prove that it is at least as hard to crack as a lattice problem.

A popular approach to lattice-based cryptography is called learning with errors (LWE), which forms the basis for several of the NIST finalists. It was introduced in 2005 by computer scientist Oded Regev at New York University. In its simplest form, it relies on arithmetic. To create a public key, the person who wants to receive a message picks a large, secret number — the private key. They then calculate several multiples of that number and add random 'errors' to each: the resulting list of numbers is the public key.

The sender adds up these whole numbers and another number that represents the message, and sends the result.

To get the message back, all the receiver has to do is divide it by the secret key and calculate the remainder. “It’s really high-school level of mathematics,” Regev says.

The profound step was Regev’s proof in 2009 that anyone who breaks this algorithm would also be able to break the seemingly more complex lattice problem. This means that LWE has the same security as lattices, but without having to deal with multi-dimensional vectors, Goldwasser says. “It’s a great formulation, because it makes it easy to work with.” Ironically, Regev discovered LWE during an unsuccessful attempt to find a quantum algorithm that would break the lattice problem. “Sometimes failure is success,” he says.



Oded Regev introduced a branch of lattice-based cryptography called learning with errors.Credit: Oded Regev

Researchers have since worked on tackling a drawback of lattice-based systems. “Lattice-based cryptography suffers from huge public keys,” says Yu Yu, a cryptographer at Shanghai Jiao Tong University in China. Whereas the public key of a current Internet application is the size of a tweet, lattice-based encryption typically requires keys that are as large as one megabyte or more. ‘Structured lattice’ systems use what are essentially algebraic tweaks to drastically reduce the public key’s size, but that can leave them more open to attack. Today’s best algorithms have to strike a delicate balance between size and efficiency.

Quantum candidates

In 2015, the NSA’s unusually candid admission that quantum computers were a serious risk to privacy made people in policy circles pay attention to the threat of Q-day. “NSA doesn’t often talk about crypto publicly, so people noticed,” said NIST mathematician Dustin Moody in a talk at a cryptography conference last year.

Under Moody’s lead, NIST had already been working on the contest that it announced in 2016, in which it invited computer scientists to submit candidate post-quantum algorithms for public-key cryptography, releasing them for scrutiny by the research community. At the same time, NIST called for submissions of digital-signature algorithms — techniques that enable a web server to establish its identity, for example, to prevent scammers from stealing passwords. The same mathematical techniques that enable public-key exchanges usually apply to this problem, too, and current digital-signature systems are similarly vulnerable to quantum attacks.

Teams from academic laboratories and companies, with members from four dozen countries on six continents, submitted 82 algorithms, of which 65 were accepted. True to their creators’ nerd credentials, many of the algorithms’ names had Star Wars, Star Trek or Lord of the Rings themes, such as FrodoKEM, CRYSTALS-DILITHIUM or New Hope.

The algorithms are being judged by both their security and their efficiency, which includes the speed of execution and compactness of the public keys.

Any algorithms that NIST chooses to standardize will have to be royalty-free.

As soon as the algorithms were submitted, it was open season. Crypto researchers delight in breaking each other's algorithms, and after NIST's submissions were made public, several of the systems were quickly broken. "I think people had a lot of fun looking at those algorithms," says Moody.

Although NIST is a US government agency, the broader crypto community has been pitching in. "It is a worldwide effort," says Philip Lafrance, a mathematician at computer-security firm ISARA Corporation in Waterloo, Canada. This means that, at the end of the process, the surviving algorithms will have gained wide acceptance. "The world is going to basically accept the NIST standards," he says. He is part of a working group that is monitoring the NIST selection on behalf of the European Telecommunications Standards Institute, an umbrella organization for groups worldwide. "We do expect to see a lot of international adoption of the standard that we'll create," says Moody.

Still, because cryptography affects sensitive national interests, other countries are keeping a close eye — and some are cautious. "The maturity of post-quantum algorithms should not be overestimated: many aspects are still at a research state," says cryptography specialist Mélissa Rossi at the National Cybersecurity Agency of France in Paris. Nevertheless, she adds, this should not delay the adoption of post-quantum systems to strengthen current cryptography.

China is said to be planning its own selection process, to be managed by the Office of State Commercial Cryptography Administration (the agency did not respond to *Nature*'s request for comment). "The consensus among researchers in China seems to be that this competition will be an open international competition, so that the Chinese [post-quantum cryptography] standards will be of the highest international standards," says Jintai Ding, a mathematician at Tsinghua University in Beijing.

Meanwhile, an organization called the Chinese Association for Cryptologic Research has already run its own competition for post-quantum algorithms. Its results were announced in 2020, leading some researchers in other

countries to mistakenly conclude that the Chinese government had already made an official choice.

Updating systems

Of [NIST's 15 candidates](#), 9 are public-key systems and 6 are for digital signatures. Finalists include implementations of NTRU and LWE, as well as another tried-and-tested system that uses the algebra of error-correction techniques. Known as ‘code-based algorithms’, these systems store data with redundancy that makes it possible to reconstruct an original file after it has been slightly damaged by noise. In cryptography, the data-storage algorithm is the public key, and a secret key is needed to reconstruct an original message.

In the next few months, the institute will select two algorithms for each application. It will then begin to draft standards for one, while keeping the other as a reserve in case the first choice ends up being broken by an unexpected attack, quantum or otherwise.

Selecting and standardizing algorithms will not be the end of the story. “It’s certainly a solid step to bless a candidate, but as a follow-up, the Internet has to agree on how to integrate an algorithm into existing protocols,” says Nick Sullivan, an applied cryptographer at Internet-services company Cloudflare, who is based in New York City.



To crack encryption, quantum computers such as China’s Jiuzhang 2.0 will need more qubits.Credit: Chao-Yang Lu

Both Cloudflare and Google — often in cooperation — have started running real-life tests of some post-quantum algorithms by including them in some beta versions of the Chrome browser and in server software. Testing is crucial because, for Internet communications to go smoothly, it is not enough to have perfectly compatible servers and browsers. To connect them, data must also run through network devices that might block traffic that they flag as unusual because of its unfamiliar encryption protocols. (These systems can be used to prevent hacking or stop users accessing prohibited content.) Antivirus software could cause similar problems. The issues also exist “on a broader, Internet-wide scale, in some countries that keep track of what users are doing”, says Sullivan. Network-security workers refer to these issues as ‘protocol ossification’, he says; it has already complicated the transition from RSA, and might disrupt the roll-out of quantum-secure algorithms, too.

An early test in 2016 implemented New Hope — a structured version of LWE named after the original *Star Wars* movie — in a Chrome beta version, and it ran without a hitch. “This trial showed that it is usable,” says Erdem Alkım, a computer scientist now at Dokuz Eylül University in İzmir, Turkey,

who wrote some of the code as part of his thesis. “I thought it was a good result for my PhD.”

But a larger-scale experiment conducted in 2021 by Google on a different algorithm ran into some snags. Some Internet devices apparently ‘broke’ — network-security parlance for a gadget that blocks a connection when a client’s browser tries to communicate with an unusual protocol. The issue could have been that the browser’s opening message was longer than expected, because it carried a large public key. Algorithms that break the Internet in this way could be shelved until these issues are resolved.

“Sometimes you run into situations in which some network element misbehaves when you add something new,” comments Rescorla. Persuading vendors to adapt their products — something that can often be done with a simple software update — could take some nudging, he says. “This could take a while.”

Still, Rescorla is optimistic, at least when it comes to Internet browsers. Because only a small number of companies control most browsers and many servers, all that needs to happen is that they change encryption systems. “Everybody is pretty confident that once NIST and IETF specify new standards, we’ll be able to roll them out pretty quickly.”

Where the transition might be trickier is the multitude of modern connected devices, such as cars, security cameras and all kinds of ‘smart home’ machines, that suffer from protocol ossification — especially those that might have security features hardwired into their chips and that are not replaced often. “It takes five to seven years to design a vehicle, and it’s going to be on the road for a decade,” says Lafrance. “Is it still going to be secure ten years down the line?”

Either way, initial implementations will be hybrid, using post-quantum technology for added security on top of existing systems. Vadim Lyubashevsky, a computer scientist at IBM in Zurich, Switzerland, whose team has two lattice-based algorithms among the NIST finalists, says he thinks both post-quantum and current encryption methods should run together for a decade before the new algorithms are used exclusively.

If all goes to plan, the Internet will be well into its post-quantum era by the time computing enters its quantum era. This post-quantum Internet could some day be followed, confusingly, by [a quantum Internet](#) — meaning a network that uses the principles of quantum physics [to make information exchange hacker-proof](#).

Researchers estimate that to break cryptosystems, quantum computers will need to have in the order of 1,000 times more computing components (qubits) than they currently do. “There’s a very good chance that we’ll have a quantum computer that can do positive things way before they can break crypto,” says Lyubashevsky.

But that is no reason to be complacent. Fully transitioning all technology to be quantum resistant will take a minimum of five years, Rescorla says, and whenever Q-day happens, there are likely to be gadgets hidden somewhere that will still be vulnerable, he says. “Even if we were to do the best we possibly can, a real quantum computer will be incredibly disruptive.”

Nature **602**, 198-201 (2022)

doi: <https://doi.org/10.1038/d41586-022-00339-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00339-5>

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

- NEWS FEATURE
- 09 February 2022

The urine revolution: how recycling pee could help to save the world

Separating urine from the rest of sewage could mitigate some difficult environmental problems, but there are big obstacles to radically re-engineering one of the most basic aspects of life.

- [Chelsea Wald](#) ✉



Specialized toilet systems recover nitrogen and other nutrients from urine for use as fertilizers and other products.Credit: MAK/Georg Mayer/EOOS
NEXT

On Gotland, the largest island in Sweden, fresh water is scarce. At the same time, residents are battling dangerous amounts of pollution from agriculture and sewer systems that causes harmful algal blooms in the surrounding Baltic Sea. These can kill fish and make people ill.

To help solve this set of environmental challenges, the island is pinning its hopes on a single, unlikely substance that connects them: human urine.

Starting in 2021, a team of researchers began collaborating with a local company that rents out portable toilets. The goal is to collect more than 70,000 litres of urine over 3 years from waterless urinals and specialized toilets at several locations during the booming summer tourist season. The team is from the Swedish University of Agricultural Sciences (SLU) in Uppsala, which has spun off a company called Sanitation360. Using a process that the researchers developed, they are drying the urine into concrete-like chunks that they hammer into a powder and press into fertilizer pellets that fit into standard farming equipment. A local farmer uses the fertilizer to grow barley that will go to a brewery to make ale — which, after consumption, could enter the cycle all over again.

The researchers aim to take urine reuse “beyond concept and into practice” on a large scale, says Prithvi Simha, a chemical-process engineer at the SLU and Sanitation360’s chief technology officer. The aim is to provide a model that regions around the world could follow. “The ambition is that everyone, everywhere, does this practice.”



The Gotland experiment compared barley fertilized with urine (right) to plants grown without fertilizer (middle) and ones with mineral fertilizer (left). Credit: Jenna Senecal

The Gotland project is part of a wave of similar efforts worldwide to separate urine from the rest of sewage and to recycle it into products such as fertilizer. That practice, known as urine diversion, is being studied by groups in the United States, Australia, Switzerland, Ethiopia and South Africa, among other places. The efforts reach far beyond the confines of university labs. Waterless urinals connect to basement treatment systems in offices in Oregon and the Netherlands. In Paris, there are plans to install urine-diverting toilets in a 1,000-resident eco-quarter being built in the 14th district of the city. The European Space Agency is to put 80 urine-diverting toilets into its Paris headquarters, which will begin operating later this year. According to proponents of urine diversion, it could see uses in sites from temporary military outposts to refugee encampments, rich urban centres and sprawling slums.

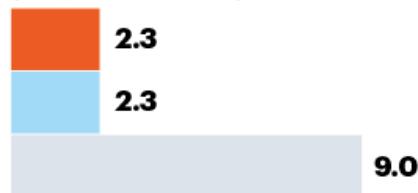
Scientists say that urine diversion would have huge environmental and public-health benefits if deployed on a large scale around the world. That's in part because urine is rich in nutrients that, instead of polluting water bodies, could go towards fertilizing crops or feed into industrial processes. According to Simha's estimates, humans produce enough urine to replace about one-quarter of current nitrogen and phosphorus fertilizers worldwide; it also contains potassium and many micronutrients (see 'What's in urine'). On top of that, not flushing urine down the drain could save vast amounts of water and reduce some of the strain on ageing and overloaded sewer systems.

WHAT'S IN URINE

A study of urban waste water globally suggests it holds enough nitrogen, phosphorus and potassium to offset more than 13% of the agricultural fertilizer demand. The value of those recovered nutrients would equal US\$13.6 billion annually.

- Potassium
- Phosphorus
- Nitrogen

Potential revenue per year
(US\$ billion, 2018)



Global fertilizer demand (%)



Nutrients in waste water (teragrams)



©nature

Source: M. Qadir *et al. Nat. Resour. Forum* **44**, 40–51 (2020)

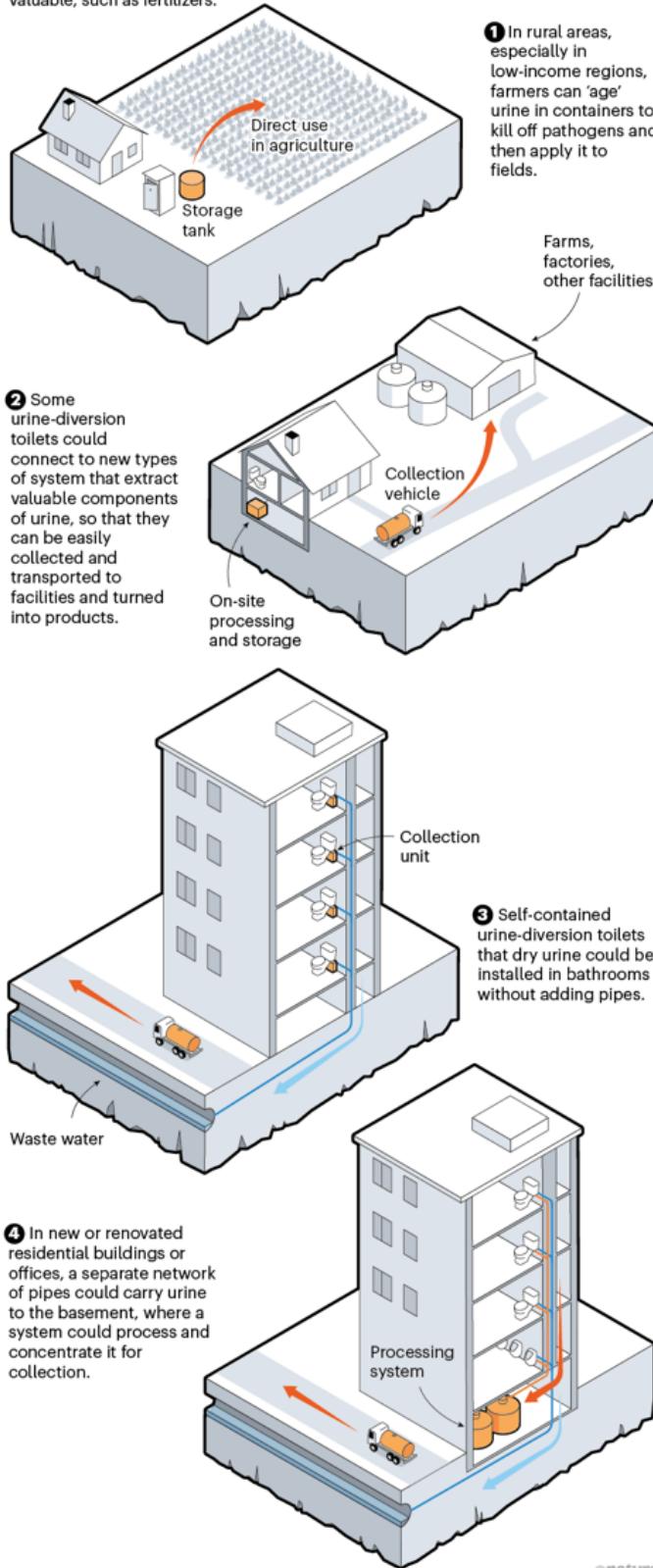
Thanks to advances in toilets and urine-treatment strategies, many components of urine diversion could soon be ready for widespread roll-out, according to experts in the field. But there are also big obstacles to radically re-engineering one of the most basic aspects of life. Researchers and companies need to solve a number of problems, from improving the design

of urine-diverting toilets to making it easier to treat urine and turn it into valuable products. This could involve chemical-treatment systems connected to individual toilets or basement devices that serve entire buildings, with pick-up and maintenance services for the resulting concentrated or solidified product (see ‘From pee to products’). Then there are broader questions of social change and acceptance, related both to varying levels of cultural taboos around human waste and to deeply entrenched conventions about industrial sewage and food systems.

Urine diversion and reuse is the type of “drastic reimagining of how we do human sanitation” that will become increasingly crucial as societies battle shortages in energy, water and raw materials for agriculture and industry, says biologist Lynn Broaddus, a sustainability consultant in Minneapolis, Minnesota, who is former president of the Water Environment Federation in Alexandria, Virginia, an association of water-quality professionals worldwide. “The fact of the matter is, it’s valuable stuff.”

FROM PEE TO PRODUCTS

Researchers are exploring various ways to make urine into something valuable, such as fertilizers.



Mixed waste

Urine used to be a valuable commodity. In the past, some societies used it for fertilizing crops, tanning leather, washing clothes and producing gunpowder. Then, in the late nineteenth and early twentieth century, the modern model of centralized sewage management arose in England and spread worldwide, ultimately leading to what has been called urine blindness.

In this model, flush toilets use water to quickly send urine, faeces and toilet paper into sewers, where it mixes with other liquids from households, industrial sources and sometimes storm run-off. At centralized treatment plants, an energy-intensive process uses microbes to clean the sewage.

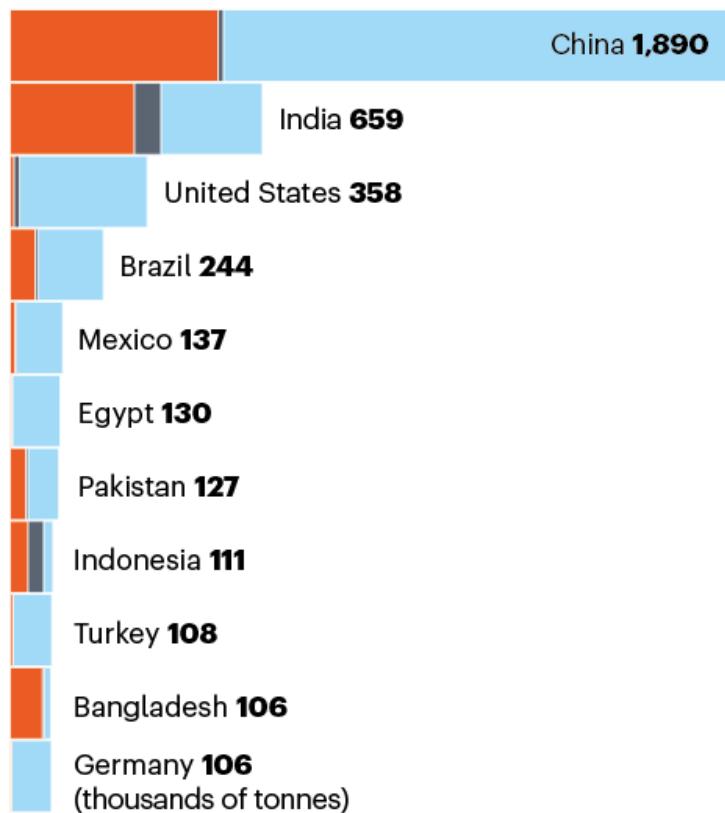
Depending on local regulations and a treatment plant's condition, the wastewater discharged from the process can still contain a lot of nitrogen and other nutrients, as well as some other contaminants. And 57% of the world's population isn't connected to centralized sewer systems at all (see 'Human sewage').

HUMAN SEWAGE

A modelling study suggests waste water adds 6.2 million tonnes of nitrogen to coastal waters around the globe. Here are the top polluters.

Sources of wastewater nitrogen that enters coastal* waters:

- No or very poor toilet facilities
- Septic tanks
- Sewers with treatment



©nature

Source: C. Tuholske *et al. PLoS ONE* **16**, e0258898 (2021).

Scientists are working on ways to make centralized systems more sustainable and less polluting, but, beginning in Sweden in the 1990s, some researchers began pushing for more fundamental change. The end-of-pipe advances are “just, you know, another evolution of the same damn thing”, says Nancy Love, an environmental engineer at the University of Michigan in Ann Arbor. Urine diversion would be “transformative”, she says. In a study¹ that modelled wastewater-management systems in three US states, she and her colleagues compared conventional wastewater systems with hypothetical ones that divert urine and use the recovered nutrients to replace synthetic fertilizers. They projected that communities with urine diversion could lower their overall greenhouse-gas emissions by up to 47%, energy consumption by up to 41%, freshwater use by about half, and nutrient pollution from the wastewater by up to 64%, depending on the technologies used.

Still, the concept has remained niche, mostly limited to off-grid locales such as northern European eco-villages, rural outhouses and development projects in low-income settings.

A lot of the lag is a result of the toilets themselves, says Tove Larsen, a chemical engineer at the Swiss Federal Institute of Aquatic Science and Technology (Eawag) in Dübendorf. First sold in the 1990s and 2000s, most urine-diverting toilets have a small basin at the front to capture the liquid — a set-up that requires careful aim. Other designs have incorporated foot-powered conveyor belts that let urine drain away while transporting the faeces to a composting vault, or sensors that operate valves to direct the urine to separate outlets.



A prototype toilet that separates urine and dries it into a powder is being tested at the head office of VA SYD, the Swedish public water and wastewater utility, in Malmö. Credit: EOOS NEXT

But in European pilot and demonstration projects, people failed to embrace their use, Larsen says, complaining that they were too unwieldy, smelly and unreliable. “We have really been stalled by this topic of toilets.”

These concerns plagued the first large-scale use of urine-diversion toilets — a project in the 2000s in South Africa’s eThekwin municipality. After apartheid, the municipality’s boundaries suddenly expanded, causing authorities to take over responsibility for some poor rural areas where there was no toilet infrastructure and little water service, says Anthony Odili, who researches sanitation governance at the University of KwaZulu-Natal in Durban.

After a cholera outbreak there in August 2000, the authorities quickly rolled out several types of sanitation that met financial and practical constraints, including about 80,000 urine-diversion dry toilets, most of which are still in use today. The urine drains below the toilet into the soil and the faeces falls into a vault, which, since 2016, the municipality has emptied every five years.

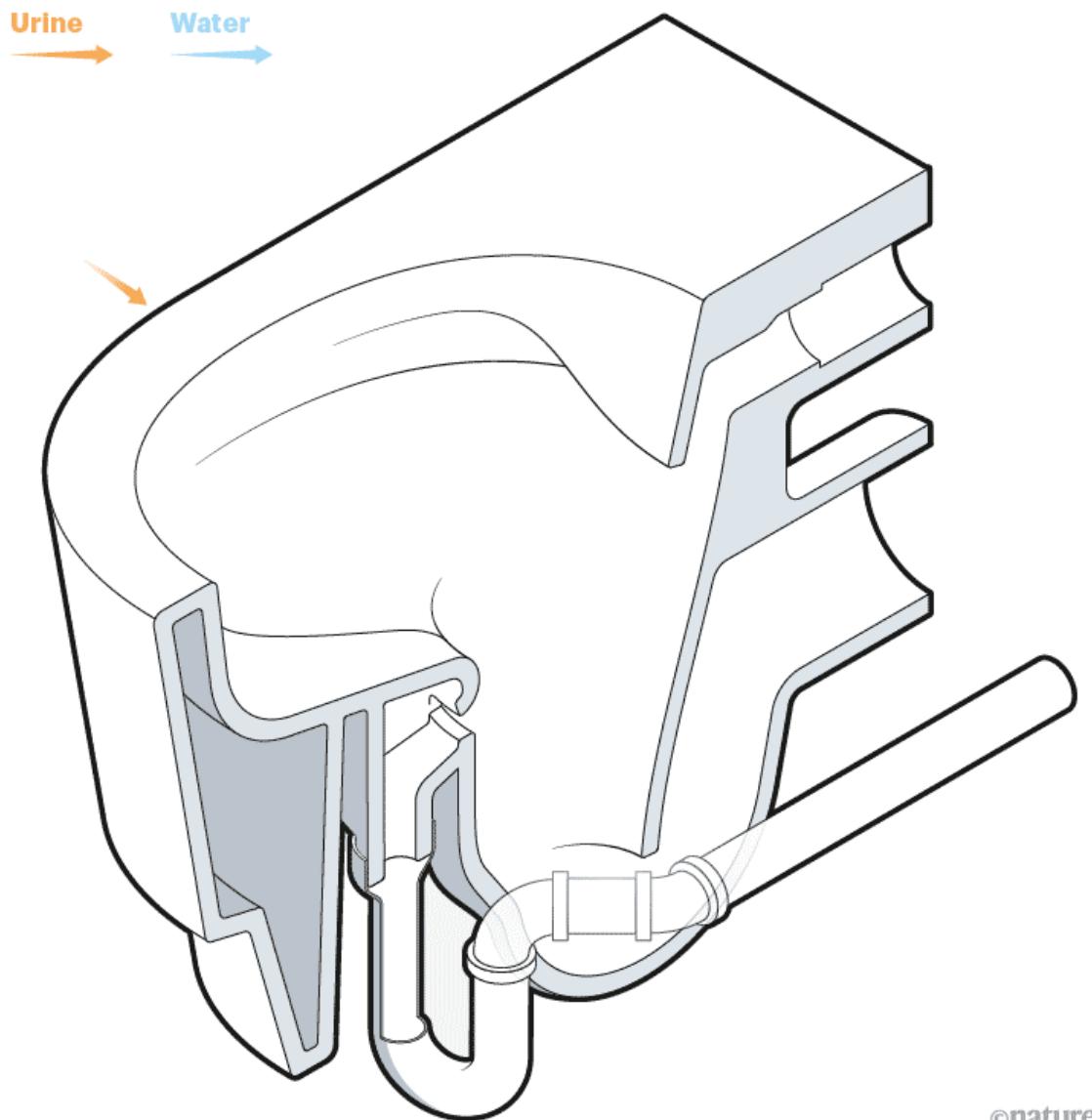
The project was successful at establishing safer sanitation in the region, Odili says. Social-science research, however, has revealed many problems with the programme. Although people felt that the toilets were better than nothing, Odili says, studies — including some he was involved in² — later found that users generally disliked them. Many had been constructed with poor materials and were awkward to use. Although such toilets should prevent bad odours in theory, urine in the eThekwin ones often entered the vaults with the faeces, causing a terrible stink. People were “not able to breathe properly”, Odili says. What’s more, the urine remains largely unused.

Ultimately, the decision to go with urine-diversion dry toilets, driven largely by public-health concerns, was top-down, and failed to take people’s preferences into account, Odili says. A 2017 study³ found that more than 95% of respondents in eThekwin aspired to the convenient, odourless flush toilets that wealthier white people use in the city — and that many have intentions to install them when their circumstances allow. In South Africa, toilets have long served as a symbol of racial disparity.

A new design, however, could represent a breakthrough for urine diversion. Led by designer Harald Gründl and in collaboration with Larsen and others, in 2017, the Austrian design firm EOOS (which has since spun off the company EOOS Next) unveiled the Urine Trap. This removes the need for users to aim, and the urine-diverting function is almost invisible (see ‘A new kind of toilet’).

A NEW KIND OF TOILET

The key to recovering nutrients from urine is to separate it from the rest of the waste stream — a process called urine diversion (UD). UD systems range from those that process urine and faeces on the spot, to those that connect to larger sewage systems. In the Urine Trap design by Austrian firm EOOS, urine runs down the front of the bowl and over a lip into a separate pipe.



©nature

Source: EOOS

It takes advantage of water's tendency to cling to surfaces (known as the teapot effect because it's like an inconveniently dribbling teapot) to direct urine down the front inner side of the toilet into a separate hole (see 'How to

recycle pee’). Developed with funding from the Bill & Melinda Gates Foundation in Seattle, Washington, which has supported a broad swathe of research into toilet innovation for low-income settings, the Urine Trap can be incorporated into everything from high-end ceramic pedestal models to plastic squat pans. LAUFEN, a manufacturer headquartered in Switzerland, is already producing one for the European market, called save!, although it is too costly for many consumers.

The University of KwaZulu-Natal and the eThekwin municipality have also been testing versions of Urine Trap toilets that divert the urine and flush the solids. This time, the research is more focused on the user. Odili is optimistic that people will prefer the new urine-diversion toilets because they smell better and are easier to use, but he points out that men would have to sit down to urinate, which is a big cultural shift. But if the toilet is “also adopted and accepted in high-income areas — people from different racial groups here — it really will help in the roll-out”, he says. “We must always put on that racial lens,” he adds, to ensure that they’re not developing something that will be seen as ‘just for Black people’ or ‘just for poor people’.

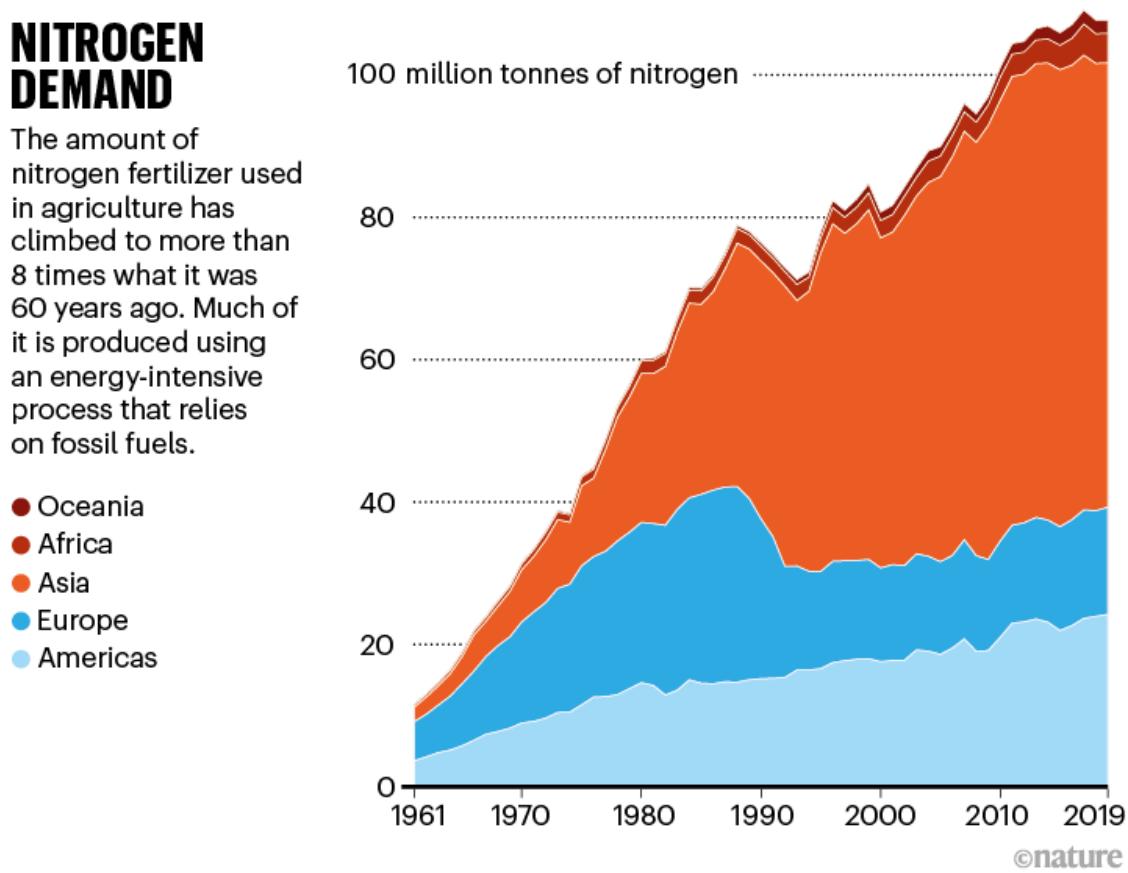
Uses for urine

Separating urine is just the first step in transforming sanitation. The next part is working out what to do with it. In rural areas, people could store it in vats to kill any pathogens and then apply it to fields. The World Health Organization provides guidelines for this practice.

But urban settings are trickier — and that’s where most urine is produced. It’s not practical to add a separate set of sewer pipes throughout a city to move urine to a central location. And because urine is about 95% water, it is too expensive to store and transport. So researchers are focusing on drying, concentrating or otherwise extracting nutrients from urine at the toilet or building level, leaving the water behind.

This isn’t easy, says Larsen. From an engineering perspective, “urine is a nasty solution”, she says. Aside from water, the largest portion is urea, a nitrogen-rich compound that bodies produce as a by-product of metabolizing

proteins. Urea by itself is useful: a synthetic version is a common nitrogen fertilizer (see ‘Nitrogen demand’). But it’s also tricky: when combined with water, the urea transforms into ammonia gas, which helps to give urine its characteristic scent. If not contained, the ammonia stinks, pollutes the air and carries valuable nitrogen away. Catalysed by the widespread enzyme urease, this reaction, called urea hydrolysis, can take microseconds, making urease one of the most efficient enzymes known⁴.



Source: FAO

Some approaches allow the hydrolysis to go ahead. Researchers at Eawag have developed an advanced process for turning hydrolysed urine into a concentrated nutrient solution. First, in a tank, microorganisms transform the volatile ammonia into non-volatile ammonium nitrate, which is a common fertilizer. Then a distiller concentrates the liquid. A spin-off company called Vuna, also in Dübendorf, is working to commercialize both the system for

use in buildings and the product, called Aurin, which has been approved in Switzerland for use on edible plants — a world first.

Others try to stop the hydrolysis reaction by quickly raising or lowering the pH of the urine, which is usually neutral when it comes out of the body. On campus at the University of Michigan, a collaboration between Love and the non-profit Rich Earth Institute in Brattleboro, Vermont, is developing a system for buildings that squirts liquid citric acid down the pipes of a urine-diverting toilet and a waterless urinal. It then concentrates the urine through repeated freezing and thawing⁵.

The SLU team doing the project on Gotland island, led by environmental engineer Björn Vinnerås, has worked out how to dry urine into a solid urea mixed with the other nutrients. The team is evaluating its latest prototype, a self-contained toilet including a built-in dryer, at the head office of the Swedish public water and wastewater utility VA SYD in Malmö.

Other methods target individual nutrients from urine. These could more easily slot into existing supply chains for fertilizers and industrial chemicals, says chemical engineer William Tarpeh, a former postdoc of Love's who is now at Stanford University in California.

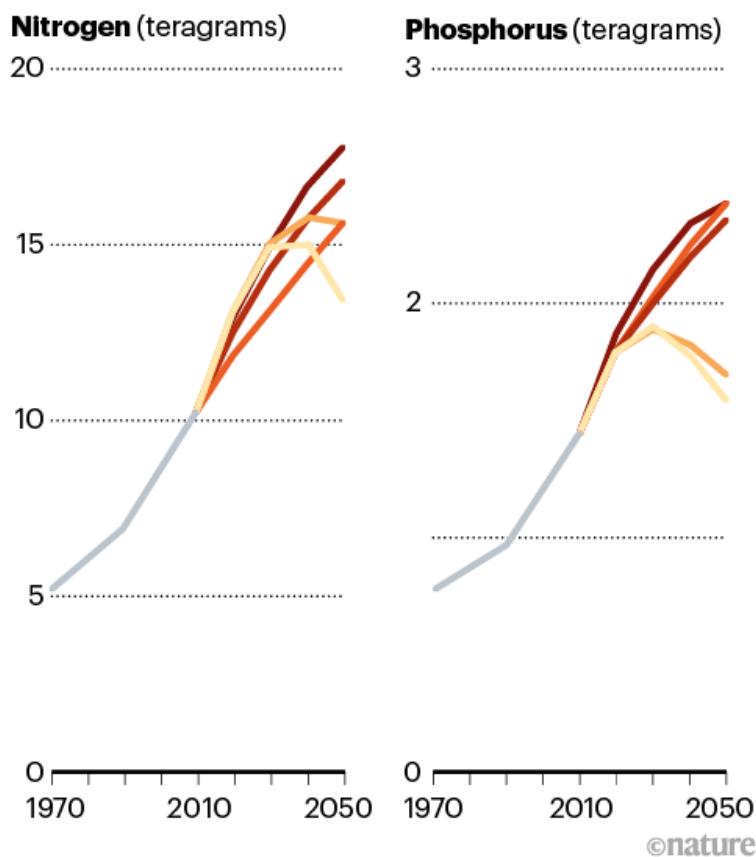
One well-established way of recovering phosphorus from hydrolysed urine is to add magnesium, which causes the precipitation of a fertilizer called struvite. And Tarpeh is experimenting with beads of adsorption materials that selectively pluck out nitrogen in the form of ammonia⁶ or phosphorus in the form of phosphate. His system uses another liquid, called a regenerant, to flow over the beads after they are spent. The regenerant carries off the nutrients and renews the beads for another round. It's a low-tech, passive method, but the commercial regenerants are environmentally damaging. His team is now trying to make ones that are cheaper and more environmentally friendly (see 'Future pollution').

FUTURE POLLUTION

The amount of nitrogen and phosphorus pollution that goes into surface waters climbs rapidly in a modelling study that explores future scenarios called Shared Socioeconomic Pathways (SSPs).

SSP scenarios:

- 1 Slow green shift
- 2 Middle of the road
- 3 Fragmented world
- 4 Inequality gaps widen
- 5 Innovation and sustainability



Source: P. J. T. M. van Puijenbroek *et al. J. Environ. Mgmt* **231**, 446–456 (2019)

Other researchers are developing ways to produce electricity by putting urine into microbial fuel cells. In Cape Town, South Africa, another team has developed a method for making an unconventional construction brick by combining urine, sand and urease-producing bacteria in a mould; these calcify into any shape without the need for firing. And the European Space Agency is eyeing astronaut urine as a resource for building habitats on the Moon.

“When I think about the big future of urine recovery and wastewater recovery, we want to be able to make as many products as possible,” Tarpeh says.

As researchers pursue a slew of ideas to turn urine into commodities, they know it’s an uphill battle, particularly with entrenched industries. Fertilizer

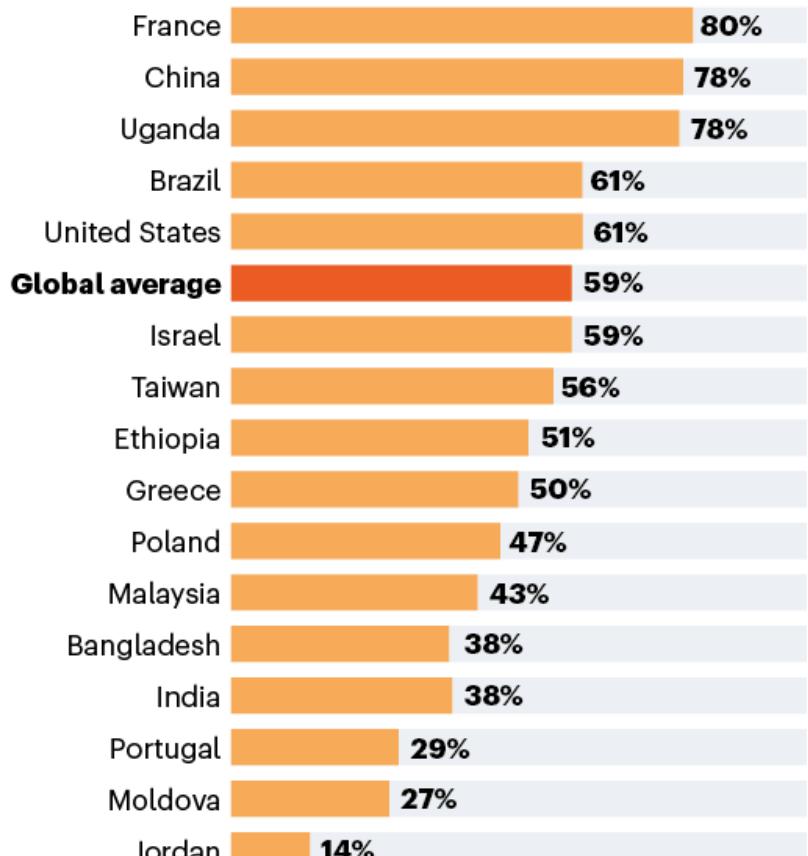
and food companies, farmers, toilet manufacturers and regulators are slow to make big changes to their practices. “There’s quite a lot of inertia,” says Simha.

At the University of California, Berkeley, for example, a research and education installation of the LAUFEN save! toilet, including a drainpipe to a storage tank on the floor below, has unexpectedly taken nearly three years and cost more than US\$50,000. That includes fees for architects, construction and complying with municipal codes, says environmental engineer Kevin Orner, now at West Virginia University in Morgantown — and it’s still not done. The lack of existing codes and regulations has caused troubles with facilities management, he says, which is why he is on a panel that is developing new codes.

Some of the inertia might be due to concerns over customer resistance, but a 2021 survey of people in 16 countries⁷ indicated that willingness to consume urine-fertilized food approached 80% in places such as France, China and Uganda (see ‘Will people eat it?’).

WILL PEOPLE EAT IT?

In a survey of more than 3,700 people at universities in 16 countries, 59% said they would eat food fertilized by urine, but there was a wide range of acceptance.



©nature

Source: Ref. 7

Pam Elardo, who leads the Bureau of Wastewater Treatment as a deputy commissioner in the New York City Department of Environmental Protection, says she supports innovations such as urine diversion, because further reducing pollution and recovering resources are key goals for her utility. The most practical and cost-effective approach to urine diversion for a city such as New York, she foresees, would be off-grid systems for renovated or new buildings, supported by maintenance and collection operations. If innovators can work that out, she says, “they should go for it”.

Given the advances, Larsen predicts that mass production and automation of urine-diversion technologies could be around the corner. And that would improve the business cases for this transformation in dealing with waste. Urine diversion “is the right technology”, she says. “It’s the only technology which can solve the problem of nutrients from households in a reasonable time. But people have to dare.”

Nature **602**, 202–206 (2022)

doi: <https://doi.org/10.1038/d41586-022-00338-6>

References

1. Hilton, S. P., Keoleian, G. A., Daigger, G. T., Zhou, B. & Love, N. G. *Environ. Sci. Technol.* **55**, 593–603 (2021).
2. Sutherland, C. *et al. Perceptions on Emptying of Urine Diverting Dehydration Toilets. Phase 2: Post eThekweni Municipality UDDT Emptying Programme* (Univ. KwaZulu-Natal, 2018).
3. Mkhize, N., Taylor, M., Udert, K. M., Gounden, T. G. & Buckley, C. A. *J. Water Sanit. Hyg. Dev.* **7**, 111–120 (2017).
4. Mazzei, L., Cianci, M., Benini, S. & Ciurli, S. *Angew. Chem. Int. Edn Engl.* **58**, 7415–7419 (2019).
5. Noe-Hays, A., Homeyer, R. J., Davis, A. P. & Love, N. G. *ACS EST Engg.* <https://doi.org/10.1021/acsestengg.1c00271> (2021).
6. Clark, B. & Tarpeh, W. A. *Chem. Eur. J.* **26**, 10099–10112 (2020).
7. Simha, P. *et al. Sci. Total Environ.* **765**, 144438 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00338-6>

Opinion

- [**Reboot biomedical R&D in the global public interest**](#) [09 February 2022]
Comment • Inequitable access to the fruits of research during the COVID-19 pandemic highlights the urgency — and feasibility — of overhauling the R&D system.
- [**Keep mountain gorillas free from pandemic virus**](#) [08 February 2022]
Correspondence •
- [**New Delhi: air-quality warning system cuts peak pollution**](#) [08 February 2022]
Correspondence •
- [**Expand diversity definitions beyond their Western perspective**](#) [08 February 2022]
Correspondence •
- [**Female scientists in Occupied Palestinian Territories call for global support**](#) [08 February 2022]
Correspondence •

- COMMENT
- 09 February 2022
- Update [10 February 2022](#)

Reboot biomedical R&D in the global public interest

Inequitable access to the fruits of research during the COVID-19 pandemic highlights the urgency — and feasibility — of overhauling the R&D system.

- [Soumya Swaminathan](#)⁰,
- [Bernard Pécoul](#)¹,
- [Hisham Abdulla](#)²,
- [Christos Christou](#)³,
- [Glenda Gray](#)⁴,
- [Carel IJsselmuiden](#)⁵,
- [Marie Paule Kieny](#)⁶,
- [Mariana Mazzucato](#)⁷,
- [Veronika von Messling](#)⁸,
- [Bernhards Ogutu](#)⁹,
- [John Reeder](#)¹⁰,
- [John-Arne Røttingen](#)¹¹,
- [Renu Swarup](#)¹²,
- [Marcel Tanner](#)¹³,
- [Nísia Trindade Lima](#)¹⁴,
- [Michelle Childs](#)¹⁵,
- [Alex Harris](#)¹⁶,
- [Els Torreele](#)¹⁷ &
- [Suerie Moon](#)¹⁸



Workers in south India make low-cost, generic HIV drugs that will expand access to therapies. Credit: Sara Hylton/Bloomberg/Getty

COVID-19 diagnostics, therapeutics and vaccines are powerful reminders: health technologies can help to shape the way in which societies control disease. Challenges in ensuring global, equitable access to these fruits of biomedical research and development (R&D) during the COVID-19 pandemic have highlighted the urgency of reorienting the system towards the public interest. The first step is a clearer articulation of what R&D in the global public interest is. That is what we seek to do here.

There are four major concerns about biomedical R&D, despite its impressive technological advances amid profound transformations in how knowledge is generated and used. The first is the lack of medicines in areas where market incentives are inadequate to attract private investment, such as for neglected diseases of poverty, bacterial infections and emerging infectious diseases¹. Second is the slow pace of progress in some areas, such as Alzheimer's disease². Third is the risk of harm, such as adverse drug reactions. The final concern is restricted access to technologies, caused by high prices, insufficient production or inadequate supply³.

These concerns pre-date the emergence of the coronavirus SARS-CoV-2, but the pandemic has underscored the urgency of addressing them. That requires looking beyond just one country or sector.

Biomedical R&D is increasingly global. Historically, it was concentrated in the advanced industrialized countries (excluding traditional medicine). Today, there is rapid growth in low- and middle-income countries (LMICs) in capacity, investment and networks. And, as medicines markets have globalized, people on every continent pay for health technologies – either out of their own pockets or through public and private medical insurance.

Both public and private interests and investments drive R&D. Research is conducted by public laboratories, universities, private firms (small, medium and multinational), non-profit organizations and health-care facilities (public and private). It is funded by taxpayers, philanthropic foundations, private investors, companies and patients. And it is shaped by public policies and

agencies, such as those for intellectual property (IP), regulatory standards, procurement, treatment guidelines and reimbursement.

All of these actors can and should reorient the biomedical R&D system to better serve the global public interest (see ‘Checklist for R&D in the global public interest’). Concretely, that means answering three questions: why do R&D? How should it be done, and for whom?

Checklist for R&D in the global public interest

Citizens, researchers, governments, intergovernmental organizations, regulators, funders, industry and universities are all stakeholders in public-interest research and development (R&D). They must collaborate to:

- Prioritize public-health needs through structured, inclusive, transparent and informed processes.
- Require that R&D is ethically conducted and scientifically sound.
- Mandate, incentivize and facilitate rapid, open sharing of inputs, processes and outputs.
- Invest in the long term to strengthen scientific, technological and regulatory capacity across all countries.
- Provide timely access to health technologies that are safe, efficacious and offer therapeutic advances.
- Ensure R&D meets the needs of subpopulations such as children, older people and those who might become pregnant.
- Recognize all contributions fairly.
- Share all benefits equitably.
- Build affordability, availability and suitability into the R&D process.

Why? Priorities

R&D should respond to priority health needs — such as for [new antibiotics](#) — as well as advancing knowledge or responding to intellectual curiosity. However, there is still a lack of systematic approaches to setting R&D priorities. Without them, the market will decide on the basis of the greatest financial return and the lowest risk. Well-organized interest groups will drown out less-powerful voices. The result? Of more than [56,000 candidate products currently under development](#), 57% are for cancerous tumours. Only 0.5% are for the [neglected tropical diseases](#) that affect nearly 2 billion people.

Investigator-driven research is the norm for early-stage discovery. But history has demonstrated the value of strategic efforts to translate science into solutions that ultimately affect people's health. For example, during the Second World War, governments drove the successful mass production of penicillin and development of malaria drugs⁴. Such 'mission-oriented' approaches require public stewardship⁵.

Because setting priorities involves value judgements, it must not be done by technical experts alone. Instead, priorities need to be set through processes that are transparent, adaptable and inclusive⁶. These can and must engage citizens, and account for the needs of patients and disadvantaged groups⁷.

Most R&D priorities transcend borders. Yet priority-setting arrangements across countries are inadequate. Countries that have the greatest burdens of neglected diseases, for example, have had little input. Instead, priorities are largely funded and decided by high-income country donors⁸. Furthermore, priority-setting requires 'gap analyses' of existing technologies and ongoing R&D. Such information, often opaque and fragmented, requires significant resources to gather and interpret.

Better priority-setting is possible. For example, during the 2015–16 Zika epidemic in Brazil, local associations of mothers of children affected by congenital Zika virus syndrome participated directly in setting national research priorities^{9,10}. The World Health Organization (WHO) Global Observatory on Health R&D collates and publishes online, open-access analyses of ongoing R&D for all therapeutic areas (see go.nature.com/3hnxxhi).

During the COVID-19 pandemic, the WHO has regularly convened funders and scientists to [identify global R&D priorities and track their progress](#). To address priorities that are specific to LMICs, 77 research organizations spanning all world regions launched the [COVID-19 Clinical Research Coalition](#). In their joint [2019 Global Action Plan for Healthy Lives and Well-being for All](#), United Nations agencies committed to supporting countries that wish to develop locally owned research agendas (see go.nature.com/3jsntnc).

But there is much more that public authorities and research funders can do to make informed, inclusive priority-setting processes the norm.

How? Ethical, sound, open, fair

All biomedical R&D must be ethical and scientifically sound. [Authoritative, international ethics guidelines](#) already exist for research involving human participants. Practices do not always meet these standards, however, nor is monitoring reliable. Increased outsourcing to contract-research organizations and the globalization of clinical trials requires close oversight to manage risks¹¹. This is particularly important when authorities have limited experience with new technologies or research regulation¹². COVID-19 has further demonstrated the value of regulators collaborating internationally to pool information and expertise; a good example is the [African Vaccine Regulatory Forum](#).

Duplicative research also raises ethical questions. Careful replication is essential for sound science. But duplication that is not part of validation exposes human participants to avoidable risks; it wastes funds, time and human resources. Open science improves efficiency and accelerates scientific progress. How? By the timely sharing of research inputs (such as specimens, compound libraries and data sets with appropriate protections), processes (such as protocols, trial designs and cost data) and outputs (including trial results and publications).

Current arrangements are inadequate for ensuring such openness, however, despite widespread recognition of its value. For example, the majority of clinical-trial outcomes are not reported on time¹³. This is despite steady

progress following clear governmental and funder policies that require all trials to be registered and their results made public¹⁴.

COVID-19 has prompted important steps forward. One is the publication of vaccine trial protocols¹⁵ and large-scale collaboration and data sharing through the WHO Solidarity and UK RECOVERY therapeutics trials¹⁶. Another is the huge increase in open sharing of genomic sequencing data on SARS-CoV-2 that enables scientists to track how, where and when the virus is changing¹⁷.

Significant changes in rules and incentives are needed to secure the rapid, open sharing of inputs, processes and outputs¹⁸. For example, IP rights can be too upstream (limiting research methods), too wide (for strategic reasons) and too strong (hard to license)¹⁹. It is in the public interest to maximize freedom to innovate, access to knowledge and follow-on research. But these require the stewardship of IP throughout the R&D process. Various initiatives to license, pool or innovate without IP demonstrate what such stewardship can look like. It is crucial to govern IP to maximize the societal benefits of knowledge, not merely to generate new inventions²⁰.



Women in Bangladesh hospitalized with dengue fever, a neglected tropical disease.Credit: Md. Mir Hossen Roney/Polaris/eyevine

Furthermore, those contributing to value creation must get fair recognition and share in its benefits²¹. Such contributions are broader than is often recognized. They can include funding, scientific expertise, infrastructure, formalized IP and traditional knowledge. Individuals or communities provide data, samples and participate in trials.

Consider these inequities. LMICs that hosted COVID-19 vaccine trials received fewer doses per capita than did high-income countries²². South African researchers publicly shared genomic sequencing data on the Omicron variant of SARS-CoV-2 through the GISAID database, enabling product developers to jump-start potential adaptations to pre-existing drugs and vaccines, but without guarantees that those products will be accessible to South Africa.

Change is feasible, albeit arduous. The 2011 WHO Pandemic Influenza Preparedness Framework painstakingly crafted a political bargain between public and private actors. It commits governments to share influenza samples openly in exchange for industry guarantees to provide certain volumes of the vaccines or other products that firms develop using the samples²³. Under the framework, countries have shared more than 1,300 samples, and the WHO has secured legally binding commitments from industry for 420 million vaccine doses in the event of an influenza pandemic²⁴. Another notable effort is the Council on Health Research for Development's [Research Fairness Initiative](#). This self-assessment tool enables research partners to examine and improve their benefit-sharing arrangements.

COVID-19 has demonstrated the feasibility of more-open R&D and wider sharing of its benefits. The GISAID platform makes genomic data openly available, while protecting certain rights of contributors to seek benefits derived from the data shared²⁵. The drug firm AstraZeneca committed to transfer technology and forgo profit from sales of the COVID-19 vaccine it jointly developed with the University of Oxford, UK, for the duration of the pandemic. The WHO created the COVID-19 Technology Access Pool, a platform for technology holders to share IP, knowledge and data with

potential product manufacturers. The UN-backed [Medicines Patent Pool](#) has negotiated licences with the drug makers Pfizer and Merck for COVID-19 treatments to be sold as low-cost generic drugs in 95 and 105 LMICs, respectively; a number of middle-income countries are excluded, however. The COVID-19 Moonshot, a “spontaneous, open, global, Twitter-fuelled collaboration” to discover and develop therapeutics for COVID-19 without patenting them, is moving towards identifying clinical-trial candidates this year^{[26](#)}.

Despite all this, there are clearly global inequalities in access to COVID-19 countermeasures. These demonstrate that much broader uptake of open science and benefit sharing are still needed, as is more learning-by-doing and regular exchange.

Finally, as COVID-19 has shown, investment in scientific, regulatory and technological capacity must be a higher political priority in countries at all levels of development (see go.nature.com/34trtjv). North–south initiatives for capacity-strengthening and technology transfer have operated over many years. Yet concerns remain about power disparities and measurable impact^{[27](#)}. An evaluation of the Wellcome-funded African Institutions Initiative, a network of 7 research capacity-building consortia across 18 African countries, found benefits in a more nationally driven, network-based approach^{[28](#)}. A group of partners has established a WHO technology-transfer hub for messenger RNA vaccines in South Africa as a multilateral, transparent mechanism to build capacity to use new technologies in LMICs; other regions have expressed interest in doing the same. Perhaps the pandemic will bolster national willingness to make the sustained investments required for adequate institutional capacity.

For whom?

Health technologies need to be safe, efficacious and high quality. Many national and international standards and institutions exist to rigorously assess the safety and efficacy of new technologies, regulate manufacturing quality and ensure adequate tracking of adverse events after a technology reaches the market. Yet significant debate remains on how exactly to carry out these duties^{[12](#)}.

Regulators need to collaborate more if technologies are to improve health quickly across different global contexts. This includes finding ways to share or accept data across borders to reduce delays. Requirements to conduct clinical trials domestically — despite relatively low numbers of cases — delayed regulatory approval of COVID-19 vaccines in Japan before it hosted the summer Olympic Games in 2021²⁹.

A review of regulations for accelerated regulatory approval of products in emergencies found more than 50 pathways in 24 countries³⁰. Important advances have been made, such as the WHO initiative to assess health technologies collaboratively with national regulators to speed up decision-making (see go.nature.com/3mgek7d). Regional efforts include the creation of the European Medicines Agency in 1995 and of the African Medicines Agency in 2021. Further development is needed.

Currently, regulators do not always require evidence of a therapeutic advance before granting approval, nor are sufficient longer-term studies necessarily conducted after a product is marketed³¹. Instead, they should require concrete improvements, such as increased efficacy, reduced toxicity, fewer adverse reactions or improved patient adherence. And they should ensure that interventions address subpopulations such as children, older people or those who might become pregnant.

To address these challenges, the use of health-technology assessment has been expanding worldwide: it often involves conducting cost–benefit analyses of competing technologies to inform policy³². For example, several European countries require evidence of therapeutic advance for a medicine to qualify for reimbursement³³. Nevertheless, more rules, incentives and financing are needed to ensure technologies benefit all groups.



A physiotherapist works with people with Alzheimer's disease, an area in which R&D has progressed slowly.Credit: Gonzalo Fuentes/Reuters/Alamy

In addition, there needs to be more emphasis on how upstream R&D decisions affect affordability and availability. For products with potentially lucrative markets, such as cancer treatments, low prices have not been an objective of commercial R&D. This can place treatments totally out of reach for most³.

In a few areas, affordability and availability are core objectives³⁴. The Drugs for Neglected Diseases initiative, for instance, has specified price targets for drugs in advance in design specifications, and has licensed IP to encourage competition between manufacturers⁷. Some COVID-19 R&D projects also aim for global access, such as the COVID-19 Moonshot²⁶, the unpatented Corbevax vaccine developed at Baylor College of Medicine in Houston, Texas, and vaccines for which R&D was largely funded by the Coalition for Epidemic Preparedness Innovations (CEPI)³⁵. It is crucial to learn from these efforts.

Given the important role that public and philanthropic funders have in all areas of R&D (and ultimately as purchasers of medicines), much more can be done to build conditionalities into health innovation. As well as stipulations about IP, supply and pricing, requirements can include transparency and reinvestment of profits into research. Public-interest research funders could work together on affordability.

Finally, high-quality health technologies must be available and appropriately designed for where and when people need them. Much of this lies outside the control of R&D actors, but not all. For example, IP can be licensed to develop new products or adapt existing ones to better suit different contexts, as happened with the HIV drug dolutegravir. The US drug firm ViiV Healthcare licensed patents on dolutegravir to the Medicines Patent Pool, which enabled generic-drugs manufacturers to combine it with two other therapies in a single pill to make treatment easier for people to take. This is now the WHO-recommended first-line therapy. Technology transfer can enable multiple manufacturers to prevent shortages and secure supply, as for AstraZeneca's COVID-19 vaccine. Developers can also submit registration dossiers promptly in all countries where products are needed³⁶.

Including these considerations in research-funding models could make health technologies more widely available.

Seize the day

We recognize that reorienting the system towards the public interest is hard. First, nationalism is not easy to align with global health concerns. Taxpayer-financed funders often have industrial, economic and political objectives, as well as health and science ones, such as competitiveness, job creation or boosting exports. Competition between nations, as witnessed in the scramble for COVID-19 vaccine doses, can undermine the willingness to cooperate.

Yet international agreements can structure cooperation to meet each country's needs³⁷, as shown by the influenza framework. International agreements could require contributions of human, financial, cultural and knowledge resources to R&D efforts in return for fair recognition and access to benefits, thereby serving the self-interest of all countries³⁸.

Second, public and private interests are not always aligned. Maximizing returns on private investment is often counter to dedicating resources to unprofitable diseases or ensuring universal affordability of medicines; openly sharing data or IP can dent competitiveness.

But COVID-19 has demonstrated that public funding and stewardship can reduce the R&D costs and risks borne by the private sector³⁹ — for example, by making affordability or data transparency more feasible. Sharing knowledge or technology is more realistic, albeit not guaranteed, when it has been co-produced by public and private actors. For example, CEPI subsidized R&D and technology transfer for US-based drug firm Novavax's COVID-19 vaccine. The first regulator-approved doses of this vaccine have been produced in India, with priority supply going to developing countries through the COVAX programme. A creative combination of regulation, incentives and persuasion can align private objectives with the global public interest.

These fundamental governance challenges are not new. Many efforts have been made to address them. These often focus on specific therapeutic areas such as neglected diseases, antibiotics and pandemics, or cover groups of countries or parts of the R&D process (basic research, discovery, translation, development, regulatory review, production, pricing and distribution). Alternatively, they might focus on certain means to achieve specific ends (an international treaty, a global R&D fund, declarations on ethical research or platforms for data sharing)³⁸.

In this article, we have taken a step back and sought to articulate a more holistic vision. COVID-19 has both exposed the shortcomings of the R&D system and offered concrete examples of how we can and must reorient it to meet the global public interest. The 1948 Universal Declaration of Human Rights recognizes people's right to "share in scientific advancement and its benefits". If not now, when?

Nature **602**, 207-210 (2022)

doi: <https://doi.org/10.1038/d41586-022-00324-y>

The views expressed in this article by WHO employees S.S. and J.R. are personal and do not necessarily represent the views, decisions or policies of the World Health Organization.

Updates & Corrections

- **Update 10 February 2022:** The author details for Carel IJsselmuiden have been updated to include a second affiliation.

References

1. Pedrique, B. *et al. Lancet Glob. Health* **1**, E371–E379 (2013).
2. Bachurin, S. O., Bovina, E. V. & Ustyugov, A. A. *Med. Res. Rev.* **37**, 1186–1225 (2017).
3. Morgan, S. G., Bathula, H. S. & Moon, S. *BMJ* **368**, L4627 (2020).
4. Gross, D. & Sampat, B. *Crisis Innovation Policy from World War II to COVID-19*. NBER Working Paper 28915 (National Bureau of Economic Research, 2021).
5. Mazzucato, M. *Mission-Oriented Research & Innovation in the European Union* (European Commission, 2018).
6. Montorzi, G., De Haan, S. & IJsselmuiden, C. *Priority Setting for Research for Health: A Management Process for Countries* (Council on Health Research for Development, 2010).
7. Drugs for Neglected Diseases initiative. *15 Years of Needs-Driven Innovation for Access* (DNDi, 2019).
8. Beyeler, N., Fewer, S., Yotebieng, M. & Yamey, G. *BMJ Glob. Health* **4**, e001209 (2019).
9. Gadelha, P., Lima, N. T., Rangel, V. F. & Stabeli, R. *Policy Focus* **13**, 23–26 (2016).

10. Silva, L., Albuquerque, M. & Mayrink, M. F. *Saúde em Debate* **45**, 861–870 (2021; in Portuguese).
11. da Silva, R. E., Amato, A. A., Guilhem, D. B. & Novaes, M. R. C. G. *Int. J. Clin. Trials* **3**, 1–8 (2016).
12. Coleman, C. H. *et al. Dev. World Bioeth.* **15**, 241–247 (2015).
13. DeVito, N. J., Bacon, S. & Goldacre, B. *Lancet* **395**, 361–369 (2020).
14. Moorthy, V. S., Karam, G., Vannice, K. S. & Kieny, M. P. *PLoS Med.* **12**, e1001819 (2015).
15. Doshi, P. *BMJ* **371**, m4058 (2020).
16. Burki, T. K. *Lancet Respir. Med.* **8**, 1178–1180 (2020).
17. Swaminathan, S. *Nature* **588**, 583–585 (2020).
18. Munos, B. *Clin. Pharmacol. Ther.* **87**, 534–536 (2010).
19. Mazzoleni, R. & Nelson, R. R. *J. Econ. Issues* **32**, 1031–1052 (1998).
20. WHO Council on the Economics of Health for All. *Governing Health Innovation for the Common Good* (World Health Organization, 2021).
21. Lazonick, W. & Mazzucato, M. *Ind. Corp. Change* **22**, 1093–1128 (2013).
22. Ramachandran, R., Ross, J. S. & Miller, J. E. *JAMA Netw. Open* **4**, e2134233 (2021).
23. Fidler, D. P. & Gostin, L. O. *JAMA* **306**, 200–201 (2011).
24. World Health Organization. *Pandemic Influenza Preparedness Framework: 18-month Progress Report, 1 January 2020–30 June 2021* (WHO, 2021).
25. Shu, Y. & McCauley, J. *Euro Surveill.* **22**, 30494 (2017).

26. von Delft, F. *et al.* *Nature* **594**, 330–332 (2021).
27. Chu, K. M., Jayaraman, S., Kyamanywa, P. & Ntakiyiruta, G. *PLoS Med.* **11**, e1001612 (2014).
28. Cochrane, G. *et al.* *The African Institutions Initiative: Insights from the First Four Years* (RAND Corporation, 2014).
29. Kosaka, M., Hashimoto, T., Ozaki, A., Tanimoto, T. & Kami, M. *Lancet* **397**, 2334–2335 (2021).
30. Simpson, S., Chakrabarti, A., Robinson, D., Chirgwin, K. & Lumpkin, M. *npj Vaccines* **5**, 101 (2020).
31. Wieseler, B., McGauran, N. & Kaiser, T. *BMJ* **366**, L4340 (2019).
32. Oortwijn, W., Broos, P., Vondeling, H., Banta, D. & Todorova, L. *Int. J. Technol. Assess. Health Care* **29**, 424–434 (2013).
33. Panteli, D. *et al.* *Health Syst. Transit.* **18**, 1–122 (2016).
34. Muñoz, V., Visentin, F., Foray, D. & Gaulé, P. *Sci. Public Policy* **42**, 315–338 (2015).
35. Lurie, N., Saville, M., Hatchett, R. & Halton, J. *N. Engl. J. Med.* **382**, 1969–1973 (2020).
36. Moran, M., Guzman, J., McDonald, A., Wu, L. & Omune, B. *Registering New Drugs: The African Context* (George Inst. for International Health, 2010).
37. Keohane, R. O. *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton Univ. Press, 1984).
38. WHO Consultative Expert Working Group on Research and Development. *Financing and Coordination. Research and Development to Meet Health Needs in Developing Countries: Strengthening Global Financing and Coordination* (World Health Organization, 2012).

39. Moon, S., Alonso Ruiz, A. & Vieira, M. N. *Engl. J. Med.* **385**, 193–196 (2021).
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00324-y>.

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

- CORRESPONDENCE
- 08 February 2022

Keep mountain gorillas free from pandemic virus

- [Kirsten Gilardi](#) ORCID: <http://orcid.org/0000-0002-6088-2178>⁰ &
- [Prosper Uwingeli](#) ORCID: <http://orcid.org/0000-0002-3833-1994>¹

As conservationists responsible for the safety of endangered mountain gorillas, we and others call on governments, researchers and tourists to continue to protect great apes from SARS-CoV-2 by adhering to current best practices for disease control in these animals (see also [T. R. Gillespie and F. H. Leendertz *Nature* 579, 497; 2020](#)).

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **602**, 211 (2022)

doi: <https://doi.org/10.1038/d41586-022-00331-z>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00331-z>

- CORRESPONDENCE
- 08 February 2022

New Delhi: air-quality warning system cuts peak pollution

- [Sachin D. Ghude](#)⁰,
- [Rajesh Kumar](#)¹,
- [Gaurav Govardhan](#)²,
- [Chinmay Jena](#)³,
- [Ravi S. Nanjundiah](#)⁴ &
- [M. Rajeevan](#)⁵

A sophisticated early-warning and decision-support system is minimizing air-pollution events in and around the Indian capital of New Delhi. This system helped to cut the city's pollution peak last November by 18–22%.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 211 (2022)

doi: <https://doi.org/10.1038/d41586-022-00332-y>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00332-y>.

- CORRESPONDENCE
- 08 February 2022

Expand diversity definitions beyond their Western perspective

- [Sakshi Ghai](#) [ORCID: http://orcid.org/0000-0002-8488-0273](#) 0

The conversation on diversity, equity and inclusion (DEI) in science needs to be expanded beyond a privileged Western perspective (see [M. McNutt and L. Castillo-Page *Nature Med.* 27, 1864–1865; 2021](#)). In many low- and middle-income countries, various religious, tribal and linguistic identities continue to be stigmatized. In India, for instance, some people are still marginalized because of their caste. To open up research to such disadvantaged groups (see [go.nature.com/2gjwkkn](#)), we need a broader definition of DEI in science.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 211 (2022)

doi: <https://doi.org/10.1038/d41586-022-00330-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00330-0>

- CORRESPONDENCE
- 08 February 2022

Female scientists in Occupied Palestinian Territories call for global support

- [Rana Dajani](#) ORCID: <http://orcid.org/0000-0002-4355-9977>⁰,
- [Elham Kateeb](#) ORCID: <http://orcid.org/0000-0003-2077-3257>¹,
- [Yara Saifi](#) ORCID: <http://orcid.org/0000-0002-7384-2503>²,
- [Maha Samman](#) ORCID: <http://orcid.org/0000-0002-8496-4004>³,
- [Rosie Dutt](#) ORCID: <http://orcid.org/0000-0002-9887-9513>⁴ &
- [Farah Qaiser](#) ORCID: <http://orcid.org/0000-0003-1050-2533>⁵

As the International Day of Women and Girls in Science approaches on 11 February, we — Palestinian scientists and members of the 500 Women Scientists collective — call attention to the continuing plight of women trying to do research in the Occupied Palestinian Territories (see [Nature 563, 308–311; 2018](#)).

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 211 (2022)

doi: <https://doi.org/10.1038/d41586-022-00314-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00314-0>

Work

- **[Start-ups create career opportunities for scientists](#)** [07 February 2022]
Career Feature • As companies compete for talent, a candidate's attitude can matter more than their credentials.
- **[Why I work unpaid to keep the Yemen Geological Museum open](#)** [26 January 2022]
Career Q&A • Despite an ongoing civil war and economic crisis, museum manager Fahd Albarraq and his colleagues want Yemenis to continue visiting the museum's collection.
- **[Weeding out the issues: preparing Brazil for cannabis growing](#)** [07 February 2022]
Where I Work • Sérgio Rocha breeds fast-growing marijuana plants at Viçosa University.

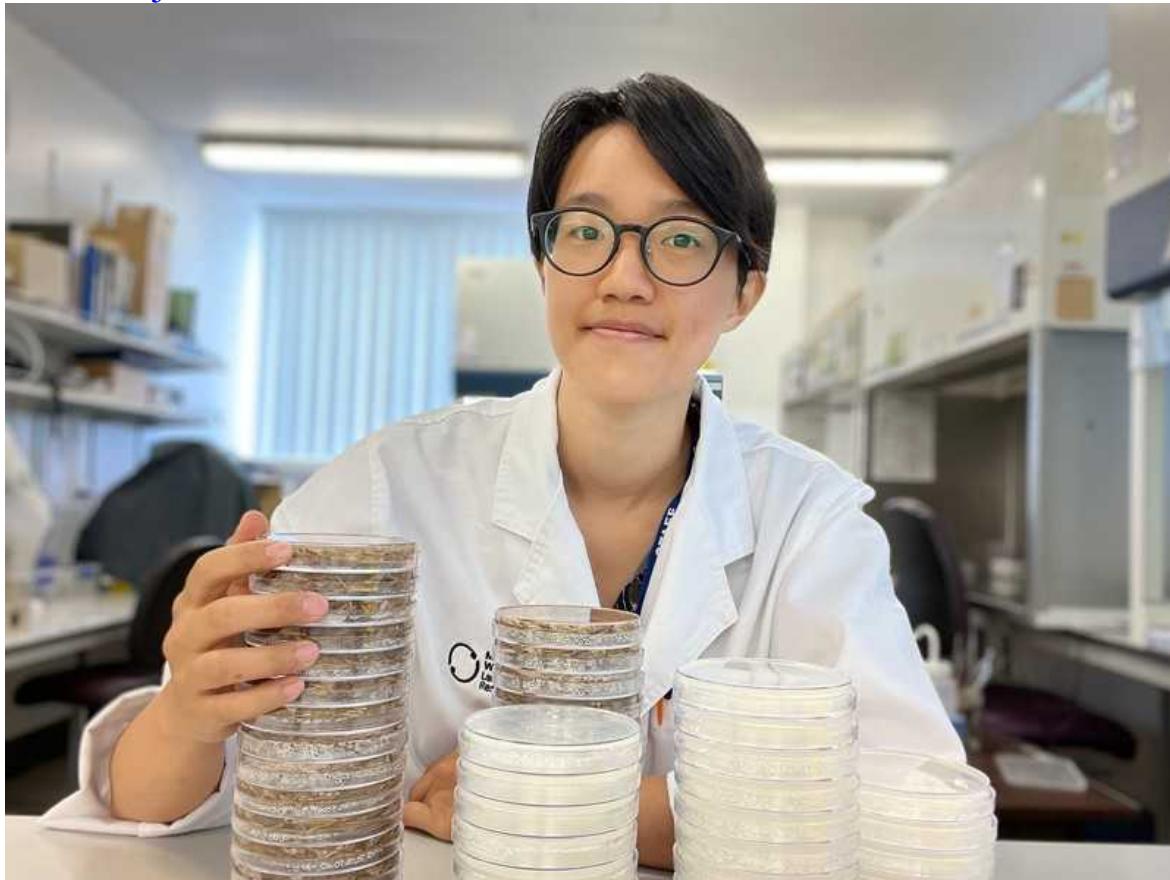
- CAREER FEATURE
- 07 February 2022

Start-ups create career opportunities for scientists

As companies compete for talent, a candidate's attitude can matter more than their credentials.

- [Chris Woolston](#) 

[Find a new job](#)



Jessica Chiang is chief scientific officer at BioFab, a start-up company in Auckland, New Zealand. Credit: Jessica Chiang

Chief science officers in industry used to fit a predictable profile: middle-aged, male and battle-hardened from many years at the bench. But the template is changing: scientific start-ups are now creating opportunities for a much wider range of researchers around the world.

Jessica Chiang, for example, was hired in 2020 as a chief scientific officer at BioFab, a start-up firm in Auckland, New Zealand, with a mission to develop biodegradable materials that could replace polystyrene. Chiang has a history of big ideas. She won a top prize at the 2017 GapSummit, a conference run by the non-profit organization Global Biotech Revolution in Washington DC, in which 100 of the world's most promising students and entrepreneurs take on some of the most pressing challenges in the biotechnology industry. But she doesn't have the standard credentials of a science officer.

As well as running research and development at BioFab, Chiang is working through the first year of her PhD programme in medical science at the University of Auckland. Through an arrangement with the university, her research at BioFab counts towards her PhD coursework, but she still has to find the time to balance life as a student and as a captain of industry. "I like to take on multiple projects and play around with different ideas," she says. "I get to work on the really science-y side of stuff, and I get to work on the commercial stuff."

Researchers looking to launch their careers have options beyond academia and big-name pharmaceutical and biotech companies. If they're willing to tolerate a bit of uncertainty and place their bets on an unproven concept, they can find a home at a newly founded company with a potentially bright future. But "a lot of scientists aren't aware of the opportunities that are out there", says Matt Krisiloff, co-founder and chief executive of Conception, a start-up in San Francisco, California, that is attempting to produce viable human eggs from stem cells with the ultimate goal of treating infertility. In his view, researchers who don't tick the usual boxes for scientific achievement can still find a good fit at a start-up, but only if they have the right attitude. "Start-ups are often willing to take bets on people based on

aptitude and enthusiasm rather than specific credentials,” says Krisiloff, who also co-founded SciFounders, a venture-capital firm in San Francisco that provides start-ups showing particular potential with US\$400,000 in exchange for a 10% equity in the firm.

Start-ups are inherently risky. An analysis of life-science companies spun off from US universities between 1980 and 2013 estimated that nearly half had failed or were on the brink of failure by 2017. Fewer than one in four were obvious successes and had been acquired by another company or were able to sell shares of stock through an initial public offering (IPO). The rest faced uncertain fates (P. [Godfrey et al. *Nature Biotechnol.* 38, 132–141; 2020](#)).

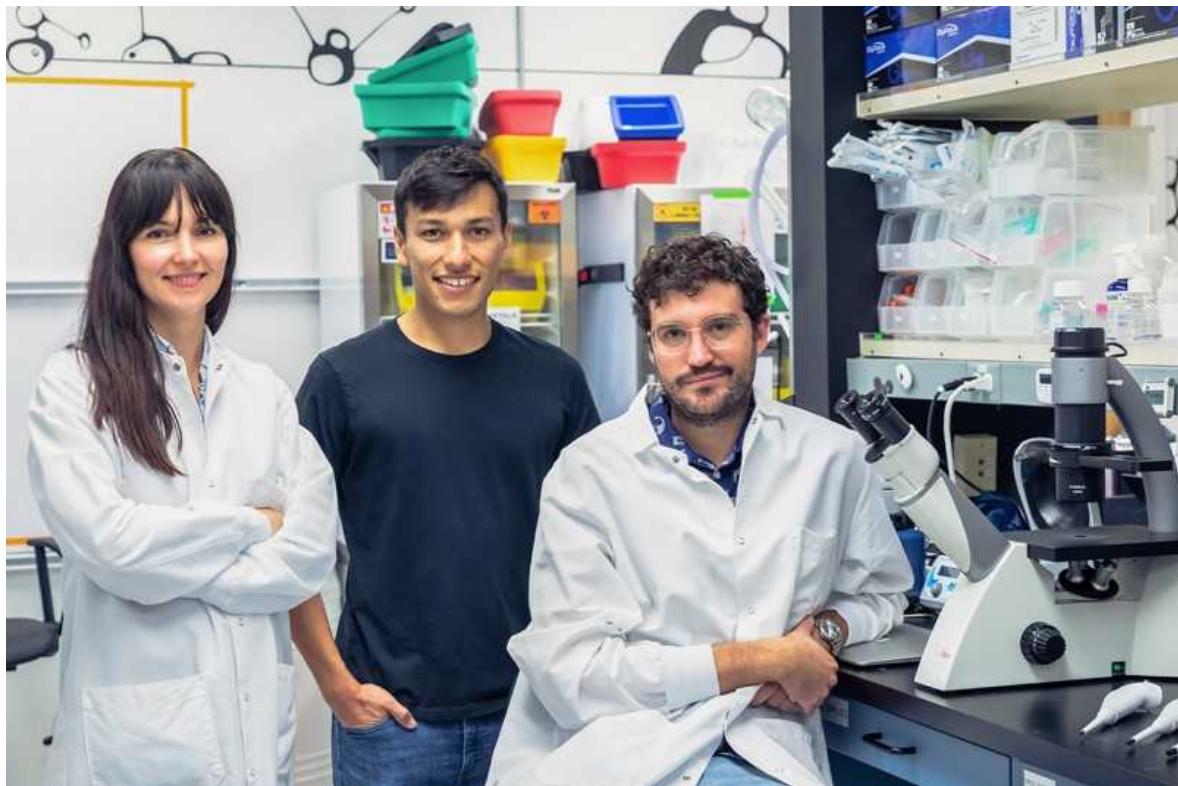
Because of this, start-ups might not suit people who have a low tolerance to taking career risks. A 2021 analysis of the salaries of Danish employees going back to 1991 suggests that relatively mature companies have generally been a safer choice. Researchers estimated that employees in start-ups earned an average of 17% less over the following 10 years than did those who joined more-established companies, partly because they tend to have periods of unemployment if the company goes under ([O. Sorenson et al. *Organ. Sci.* 32, 587–604; 2021](#)).

Desperately seeking scientists

But Krisiloff argues that joining a start-up isn’t as risky as it might seem. “Companies are getting started on a weekly basis,” he says. BDO, a global financial-services company, reports that 78 US biotech companies offered IPOs in 2020, an all-time high and a 77% increase from the previous year (see [go.nature.com/3s8yu2q](#)). “If the company you’re at doesn’t work out, you can move easily because of the shortage of good scientists,” he adds. “It’s the opposite of academia, where there are so many postdocs and so few professorships.”

Just as labour shortages are plaguing workplaces such as restaurants, factories and farms in some parts of the world, many start-up companies are struggling to fill posts, Krisiloff says. “There’s so much capital sloshing around in the world right now,” he says. “Companies that are not that far

along are raising hundreds of millions of dollars and have to figure out ways to justify that funding. They're desperately looking for people."



Matt Krisiloff (centre) with Conception co-founders Bianka Seres (left) and Pablo Hurtado González.Credit: Christopher Williams

Companies are often willing to invest in talent. Salaries vary widely from company to company and from field to field, but Krisiloff says that typical starting salaries for PhDs in the life sciences in the United States is around \$120,000 a year, including full benefits. In 2020, Labiotech.eu, a media site that covers the biotech industry in Europe, reported that senior scientists, which includes many new recruits at start-ups, can expect to earn up to €70,000 (US\$78,000) a year.

Last year, Krisiloff started [a Twitter thread](#) to help connect short-staffed start-ups with scientists looking for work. He asked start-ups to explain their mission and the type of people they were looking to hire. The initial tweet generated responses from companies with a wide variety of products, including gene therapy, microscopy and laboratory-grown meat.

Mammoth Biosciences, a biotechnology company in Brisbane, California, posted to the thread that it had “many open positions”. The company was founded in 2017 by two recent PhD graduates from Stanford University and two PhD students at the University of California, Berkeley. They were joined by biochemist and gene-editing pioneer Jennifer Doudna, who now chairs its scientific advisory board. The company, which uses CRISPR technology for applications in areas such as health care, biodefence and agriculture, now has 132 full-time employees (including 42 PhD researchers) and another 30 job openings for scientists and engineers, says chief operations officer Ted Tisch.

Open to suggestions

Réka Trón, co-founder of Multus, a London-based company that produces the media needed to grow meat in a lab, tweeted a link to current job openings, adding: “If you can’t find one that suits you, e-mail us. We might hire you!” The Multus website says that “your motivation is more important to us than perfect grades, university degrees and a complete curriculum vitae”.



Réka Trón is chief operating officer at London-based Multus. Credit: Imperial College London

Multus, incorporated in March 2020, now has 11 employees. Trón, who founded the company with two fellow students at Imperial College London and is now chief operating officer, says she is actively seeking someone who can serve as both a computational biologist and a software engineer. But she's open to hearing from someone who is willing and able to learn some skills on the job. She emphasizes that she would welcome an application from any qualified scientist who is excited about the company's mission and reducing the impact of livestock agriculture. "If someone brings a great idea or a great value to the team, we're open to the possibility," she says.

Multus is locked in competition for scientific talent with other firms, including the dozens working on lab-grown meat. Trón says the company still has to be discerning. In a smaller firm, there is less room for people who

can't work as part of the team or get along well with others. "Attitude is an extremely important part of hiring," she says.

The company was formed during the pandemic, so Trón and her team have had to rely on Zoom interviews to gauge applicants' personalities. Candidates who live nearby are encouraged to come for a tour of the lab. She says that people tend to be more conversational and less stressed when they aren't talking to a screen.

A 'help wanted' sign is definitely up at Conception. "We're always looking for stem-cell scientists and organic biologists who are interested in our mission," Krisiloff says. "We're not big on credentials. We're happy to consider someone who may not have a PhD."

Krisiloff says that scientists who apply to start-ups often have a deep supply of curiosity and a tolerance for uncertainty. For various reasons, they also tend to be eager to leave the conventional academic career path. Not only can they find more job openings in industry, but they can also have more time to focus on their work. "It can be a much more effective way of actually doing research," Krisiloff says. "Unlike in academia, you don't have everyone trying to carve out their own little niches that they have to publish around."

Big-company comparisons

In the right circumstances, a start-up can be a launching point for a stable career with ample room for promotion. Twelve years ago, immunologist Laurent Poirot left a postdoctoral position to join Cellectis, a biotech pharmaceutical firm in Paris. Poirot rose through the ranks and is now a senior vice-president of immunology in charge of a ten-person team. He enjoys mixing management responsibilities and business savvy with pure research. "The most rewarding and gratifying thing is having a [therapy] in my hands that I could see being injected into someone," Poirot says. "To be a part of that from inception to the bedside is exciting. That never happens in academia, unless you're in a lab that does its own clinical trials."

The firm, which was founded in 1999 and has about 280 employees, is now distant from its start-up roots, but Poirot says it still has plenty of untested ideas for new pharmaceutical products for anyone who wants to get in on the ground floor of something big.

Poirot says he's especially interested in hiring people who have shown when experiments or projects didn't go as planned. "Having been exposed to failure in science is something that I find tremendously valuable," he says. If an applicant says they have never faced real failure, Poirot will use the interview process to gauge their commitment to solving problems. "I like to challenge people when I talk to them. If the call turns into an ad-libbed discussion about science, that gets me excited."

It's much too early to predict the ultimate fate of BioFab, but Chiang says she plans to stay there at least until she finishes her PhD, which could take four years or so. By that time, she hopes the company will have a prototype that can generate sales. She also hopes that she's merely at the beginning of a long career in the start-up world. "I want to stay with this company and grow it, and maybe start other companies in the future," she says. "I'm an entrepreneur for life."

Nature **602**, 349-351 (2022)

doi: <https://doi.org/10.1038/d41586-022-00343-9>

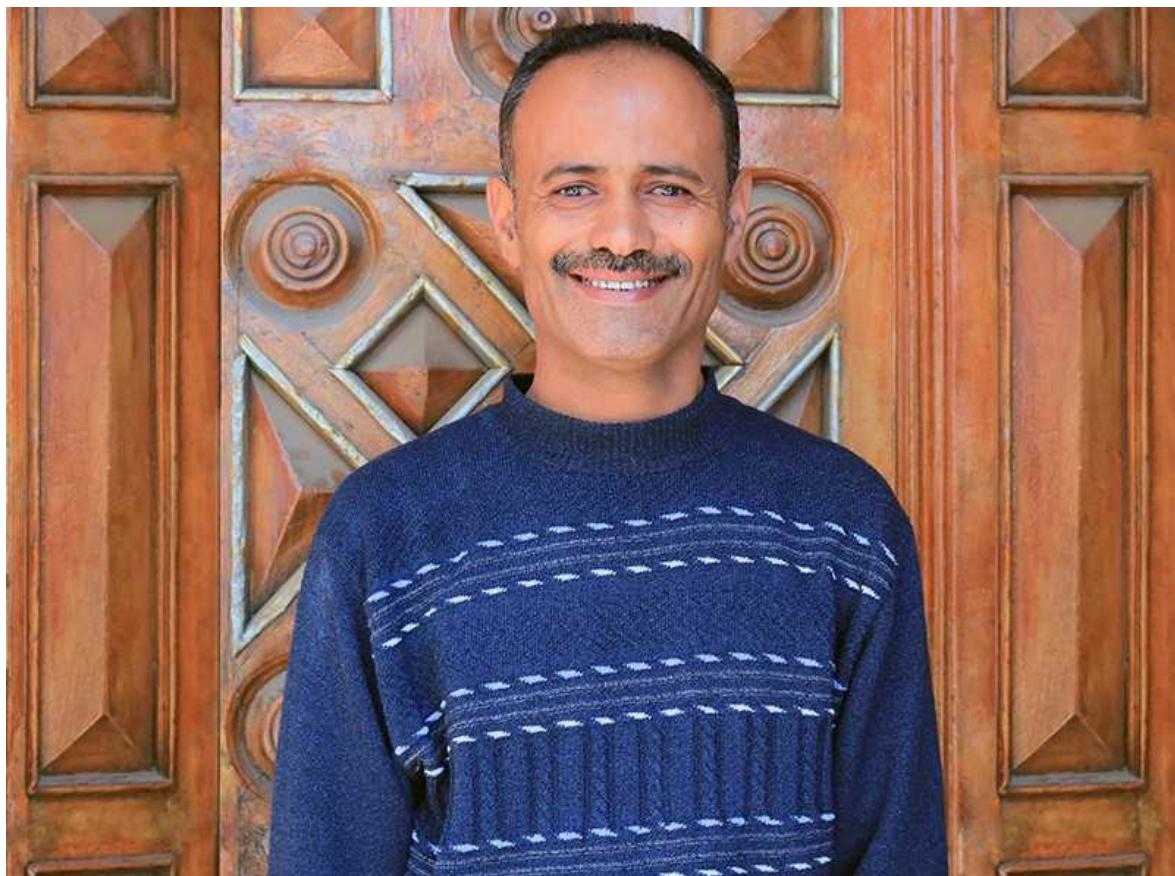
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00343-9>

- CAREER Q&A
- 26 January 2022

Why I work unpaid to keep the Yemen Geological Museum open

Despite an ongoing civil war and economic crisis, museum manager Fahd Albarraq and his colleagues want Yemenis to continue visiting the museum's collection.

- [Shihab Jamal](#) 0



Geologist Fahd Albarraq has made many sacrifices to keep the Yemen Geological Museum open. Credit: Shihab Jamal

Yemeni geologist Fahd Albarraq manages the Yemen Geological Museum in Sana'a. In 2016, not long after civil war broke out, the country's economy collapsed and the museum lost its permanent financing from the government. Albarraq and three of his colleagues reopened it, without pay, in 2017.

Tell us about the museum and your exhibits.

It's a scientific museum that was founded in 1999 and obtained official recognition from the government in 2012, giving it national museum status. I've worked here for most of my professional life. Our main focus is on geology, thanks in part to our sponsors, the Yemen Geological Survey and Mineral Resources Board (YGSMRB). And, as well as our exhibits, we have a 140-seat lecture theatre, where we screen documentary films.

During the 2000s, we added more exhibits, such as snake fossils from the Jurassic period (201 million to 145 million years ago) and fish fossils from the Silurian period (443 million to 419 million years ago). There are also fossils of many types of mollusc and echinoderm, and of plants and trees from different geological periods. Almost all of the exhibits are from Yemen. We also show types of igneous, sedimentary and metamorphic rock, such as rhyolite, shale and gneiss.

Since 2019, the YGSMRB has given volunteers US\$20 per month to cover our commuting costs. Visitors can also donate to the museum.

What are the difficulties you face?

The suspension of salaries — which happened in 2016 as a result of the war — is our biggest problem. There were 12 of us before the conflict started, and only 4 have been able to return to the museum to work on a voluntary basis.

In 2019 we also lost our government funding. This meant we couldn't pay for the museum's website, so we were unable to share any news. We now depend on our Facebook page, which is free to run. Despite the financial difficulties, admission to the museum is still free and will remain so. Before the civil war broke out, the museum averaged around 60 visitors a day. By 2021, that number had increased to more than 140 a day, thanks to a combination of social media promotion and our encouraging school trips. The war itself is also a factor: people are looking for escape and distraction in science and art.

How has the war affected you personally?

Before the war, I lived in comfort. As a geology student at Sana'a University between 1997 and 2001, I worked in a fast-food restaurant after classes. The wages were enough to live a good life because prices were low, and I could send money to my father, who lives in a Yemeni mountain village. In 2002, I got the job at the museum and was able to save money easily. I got married the following year, and a few years later bought a car.

Now, I borrow money from many people. I'm burdened by a monumental amount of debt. I have two daughters and two sons, aged from 6 to 16, and I have sold my car and my wife's jewellery to pay their school fees.

The war threatens our safety and our financial situation. Killing and destruction is everywhere. The most awful thing is the intense air raids on Sana'a, where we live. I worry that I can't do anything to make my family feel safe.

Why do you still volunteer at the museum?

This museum is an important part of my life. When a person works in a field they love, they sacrifice a lot to stay in that job. I wouldn't be comfortable with another job outside the field of earth sciences.

Working at this museum makes me feel that I am doing something worthwhile. After I finish my work at the museum, I go to my uncle's small mobile-phone shop. He gives me wages that help me meet some family

expenses. It's hard to consider it a good opportunity, but it allows me to keep working at the museum.

Nature **602**, 351 (2022)

doi: <https://doi.org/10.1038/d41586-022-00181-9>

This interview has been edited for length and clarity.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00181-9>

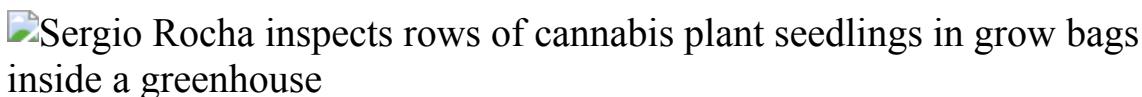
| [Section menu](#) | [Main menu](#) |

- WHERE I WORK
- 07 February 2022

Weeding out the issues: preparing Brazil for cannabis growing

Sérgio Rocha breeds fast-growing marijuana plants at Viçosa University.

- [Patrícia Maia Noronha](#) 0



Sérgio Rocha is an agronomist engineer and executive director of ADWA Cannabis in Viçosa, Credit: Washington Alves/Reuters/Alamy

In this picture, I'm trying to work out which cannabis plants grow the fastest. I'm looking for those that have started to flower early, indicating that their reproductive organs have become active. My team and I can then cross-breed them for genetic improvement.

We have a 100-square-metre greenhouse at Brazil's Federal University of Viçosa. Growing cannabis for medical purposes is legal in Brazil, but only a few projects have been authorized. ADWA Cannabis, our company, is allowed to cultivate cannabis for research purposes here on campus — but we cannot expand beyond the university. Soon, we plan to open a new greenhouse, still on campus, three times bigger than this one.

Around 2015, when I started my bachelor's degree at Viçosa, the restrictive cannabis laws prompted nationwide protests — promoted mainly by mothers of kids with disorders such as epilepsy, who might benefit from drugs made from cannabis compounds but struggle to access them. Now, the Brazilian congress is considering making cannabis cultivation for

medical and industrial purposes fully legal. Brazil would then probably attract investment from countries including Canada and the United States, which have expertise in cannabis products. We hope that Brazilian companies can also carve out a space in this market.

When my partners and I founded ADWA Cannabis in 2018, we had only four subspecies of plant. After generations of cross-breeding, we are now evaluating 38 possible new varieties. Ultimately, our goal is to create plants that are completely comfortable in Brazilian soil and weather.

Besides the genetic improvement, we focus on monitoring the cultivation process, from soil nutrition to lighting and pest control. For this, we've developed software that has attracted the interest of Uruguayan and Spanish cannabis companies. We hope to license this software to more countries where cannabis cultivation is legal.

Nature **602**, 354 (2022)

doi: <https://doi.org/10.1038/d41586-022-00344-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00344-8>

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

Research

- **[Neural networks overtake humans in Gran Turismo racing game](#)** [09 February 2022]

News & Views • Driving a racing car requires a tremendous amount of skill. Now, artificial intelligence has challenged the idea that this skill is exclusive to humans — and it might even change the way automated vehicles are designed.

- **[A hard graft problem solved for key global food crops](#)** [25 January 2022]

News & Views • Grafting has long been used to join tissues of different plants in horticulture and research. Methods have now been devised to extend the technique to plants called monocotyledons, which include major crops such as cereals and bananas.

- **[Electrons go loopy in a family of superconductors](#)** [09 February 2022]

News & Views • Measurements indicate that electrons move in loops between the atoms of an intriguing class of superconducting material. Such dynamics breaks key symmetries of the crystal lattice — suggesting the material hosts a rare state of matter.

- **[From the archive](#)** [08 February 2022]

News & Views • Nature's pages feature the discovery of a key defence response, and a collection of items from an inventor highlights developments in photography.

- **[Inflammation in the gut is encoded by neurons in the brain](#)** [10 January 2022]

News & Views • The nervous and immune systems interact in a bidirectional manner. It emerges that inflammation in the body activates brain cells that, when later reactivated, can trigger a recapitulation of the inflammatory response.

- **[A white dwarf accreting planetary material determined from X-ray observations](#)** [09 February 2022]

Article • An X-ray source is detected at the expected position of the white dwarf star G29–38, which enables the calculation of the accretion rate of planetary material without using stellar atmosphere models.

- **[Outracing champion Gran Turismo drivers with deep reinforcement learning](#)** [09 February 2022]

Article • Using the game Gran Turismo, an agent was trained with a combination of deep reinforcement learning algorithms and specialized training scenarios, demonstrating success against championship-level human racers.

- **Evidence for the association of triatomic molecules in ultracold 23Na40K + 40K mixtures** [09 February 2022]
Article • Evidence is presented for the association of triatomic molecules near the Feshbach resonance in an ultracold gas comprising a mixture of 23Na40K molecules and 40K atoms, along with an estimation of the binding energy of the triatomic molecules.
- **Real-space visualization of intrinsic magnetic fields of an antiferromagnet** [09 February 2022]
Article • Real-space visualization of the magnetic fields in antiferromagnetic haematite is achieved using atomic-resolution differential phase contrast scanning transmission electron microscopy in a magnetic-field-free environment.
- **Ferroelectric incommensurate spin crystals** [09 February 2022]
Article • Analysing the structure of a PbTiO₃ epitaxial layer sandwiched between SrRuO₃ electrodes led to observation of a topology with two periodic modulations that form an incommensurate polar crystal, providing an analogue to incommensurate spin crystals.
- **Time-reversal symmetry-breaking charge order in a kagome superconductor** [09 February 2022]
Article • An investigation of muon spin relaxation shows time-reversal symmetry-breaking charge order, intertwined with correlated superconductivity, due to orbital currents in the kagome superconductor KV₃Sb₅.
- **A highly distorted ultraelastic chemically complex Elinvar alloy** [09 February 2022]
Article • A chemically complex alloy that exhibits a high elastic strain limit and low internal friction is described; it also has an Elinvar effect (invariant elastic modulus) over a large temperature range, up to 627 °C.
- **Superionic iron alloys and their seismic velocities in Earth's inner core** [09 February 2022]
Article • Molecular dynamics simulations show that the light elements hydrogen, oxygen and carbon become highly diffusive like liquid in solid iron under the inner-core conditions, leading to a reduction in the seismic velocities.
- **A species-level timeline of mammal evolution integrating phylogenomic data** [22 December 2021]
Article • Bayesian analysis of datasets comprising genomes from multiple mammalian species can efficiently and precisely decipher their evolutionary timeline.
- **Autism genes converge on asynchronous development of shared neuron classes** [02 February 2022]

Article • Haploinsufficiency in three genes associated with risk of autism spectrum disorder—KMT5B, ARID1B and CHD8—in cell lines from multiple donors results in cell-type-specific asynchronous development of GABAergic neurons and cortical deep-layer excitatory projection neurons.

- **Cortical preparatory activity indexes learned motor memories** [26 January 2022]

Article • In rhesus monkeys, learning of a motor task is accompanied by uniform changes in preparatory activity in motor cortex that are orthogonal to the force-predictive neural state subspace.

- **Monocotyledonous plants graft at the embryonic root-shoot interface** [22 December 2021]

Article • Intra- and inter-specific grafting is possible in most orders of monocotyledonous plants, and this process could be used to combat diseases that affect crops, such as Panama disease in bananas.

- **Morphogen gradient scaling by recycling of intracellular Dpp** [22 December 2021]

Article • A mechanism involving intracellular recycling of the morphogen Decapentaplegic (Dpp) underlies the scaling of the Dpp gradient in the *Drosophila* wing disc, and this is modulated by the extracellular factors Pentagone and Dally.

- **The N501Y spike substitution enhances SARS-CoV-2 infection and transmission** [24 November 2021]

Article • Experiments in a hamster model of COVID-19 and human airway epithelial cells show that the spike N501Y mutation is the major determinant of increased fitness of the B.1.1.7 Alpha variant of SARS-CoV-2.

- **Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation** [25 November 2021]

Article • The P681R mutation in the spike protein renders the Delta variant more pathogenic than prototypic SARS-CoV-2 in infected hamsters, and facilitates spike protein cleavage and enhances viral fusogenicity.

- **Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta** [22 December 2021]

Article • The Alpha variant of SARS-CoV-2 outcompetes progenitor SARS-CoV-2 in upper respiratory tract replication competition *in vivo*.

- **Broadly neutralizing antibodies target a haemagglutinin anchor epitope** [23 December 2021]

Article • A distinct class of broadly neutralizing antibodies to the influenza virus target a membrane-proximal anchor epitope of the haemagglutinin stalk domain.

- **Local and systemic responses to SARS-CoV-2 infection in children and adults** [22 December 2021]

Article • Mechanisms explaining the milder clinical syndrome that is observed in children with SARS-CoV-2 infection.

- **Activation mechanism of PINK1** [21 December 2021]

Article • Unphosphorylated PINK1 of Pediculus humanus corporis forms a dimerized state before undergoing trans-autophosphorylation, and phosphorylated PINK1 undergoes a conformational change in the N-lobe to produce its phosphorylated, ubiquitin-binding state.

- **Crystallographic snapshots of a B12-dependent radical SAM methyltransferase** [02 February 2022]

Article • Structural and spectroscopic studies show how a B12-dependent radical SAM enzyme catalyses unique and challenging alkylation chemistry, including protein post-translational modification required for methane biosynthesis.

- **Structure of a B12-dependent radical SAM enzyme in carbapenem biosynthesis** [02 February 2022]

Article • X-ray crystal structures of TokK, a cobalamin- or B12-dependent radical SAM methylase, provide insight into how these enzymes use sequential radical-mediated methylations to assemble the C6 side chain of carbapenem antibiotics.

- NEWS AND VIEWS
- 09 February 2022

Neural networks overtake humans in *Gran Turismo* racing game

Driving a racing car requires a tremendous amount of skill. Now, artificial intelligence has challenged the idea that this skill is exclusive to humans — and it might even change the way automated vehicles are designed.

- [J. Christian Gerdes](#) 0

A modern Formula 1 race is a breathtaking display of engineering precision. Yet the popularity of the sport arguably has less to do with the performance of the cars than with the skill and daring displayed by the drivers as they push those cars to the limit. Success on the race track has been a celebrated human achievement for more than a century. Will it now become a similar triumph for artificial intelligence (AI)? [Writing in Nature](#), Wurman *et al.*¹ take a step in this direction by introducing Gran Turismo (GT) Sophy, a neural-network driver capable of outperforming the best human players of the video game *Gran Turismo*.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 213-214 (2022)

doi: <https://doi.org/10.1038/d41586-022-00304-2>

References

1. Wurman, P. R. *et al.* *Nature* **602**, 223–228 (2022).

2. Spielberg, N. A., Brown, M., Kapania, N. R., Kegelman, J. C. & Gerdes, J. C. *Sci. Robot.* **4**, eaaw1975 (2019).
3. Fujiyoshi, H., Hirakawa, T. & Yamashita, T. *IATSS Res.* **43**, 244–252 (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00304-2>

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

- NEWS AND VIEWS
- 25 January 2022

A hard graft problem solved for key global food crops

Grafting has long been used to join tissues of different plants in horticulture and research. Methods have now been devised to extend the technique to plants called monocotyledons, which include major crops such as cereals and bananas.

- [Colin Turnbull](#) ⁰ &
- [Sean Carrington](#) ¹

The technique of grafting together the shoot of one plant and the roots of another is immensely beneficial in a variety of contexts. However, efforts to use this approach have long failed for certain key crops. [Writing in Nature](#), Reeves *et al.*¹ report success in developing a grafting method that can be used for plants called monocotyledons, or monocots.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 214-215 (2022)

doi: <https://doi.org/10.1038/d41586-022-00050-5>

References

1. Reeves, G. *et al.* *Nature* **602**, 280–286 (2022).
2. Gautier, A. T. *et al.* *J. Exp. Bot.* **70**, 747–755 (2019).

3. Turnbull, C. G. N. in *Plant Developmental Biology: Methods and Protocols* (eds Hennig, L. & Köhler, C.) 11–26 (Humana, 2010).
4. Runo, S. & Kuria, E. K. *PLoS Pathog.* **14**, e1006731 (2018).
5. Melnyk, C. W. *et al.* *Proc. Natl Acad. Sci. USA* **115**, E2447–E2456 (2018).
6. Viljoen, A., Ma, L.-J. & Molina, A. B. in *Emerging Plant Diseases and Global Food Security* (eds Ristaino, J. B. & Records, A.) Ch. 8, 159–184 (APS, 2020).
7. Powell, K. S., Cooper, P. D. & Forneck, A. *Adv. Insect Physiol.* **45**, 159–218 (2013).
8. Warschefsky, E. J. *et al.* *Trends Plant Sci.* **21**, 418–437 (2016).
9. Gomez-Roldan, V. *et al.* *Nature* **455**, 189–194 (2008).
10. Umehara, M. *et al.* *Nature* **455**, 195–200 (2008).
11. Kondhare, K. R., Patil, N. S. & Banerjee, A. K. *J. Exp. Bot.* **72**, 4218–4236 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00050-5>

- NEWS AND VIEWS
- 09 February 2022

Electrons go loopy in a family of superconductors

Measurements indicate that electrons move in loops between the atoms of an intriguing class of superconducting material. Such dynamics breaks key symmetries of the crystal lattice — suggesting the material hosts a rare state of matter.

- [Morten H. Christensen](#) ⁰ &
- [Turan Birol](#) ¹

The exotic properties of materials can often be traced to the collective behaviours of their electrons. These behaviours give rise to different electronic phases of matter, which physicists identify by studying symmetries of the material: a broken symmetry signals a new electronic phase. One such phase is a charge density wave, in which electrons organize into a pattern of high and low concentrations of electric charge. In [a paper in *Nature*](#), Mielke *et al.*¹ report that a material belonging to a group of compounds known as kagome superconductors^{2,3} exhibits a charge density wave with electron dynamics that break time-reversal symmetry — the electronic state is different depending on whether time is moving forwards or backwards. [Writing online in *Nature*](#), Nie *et al.*⁴ show that a charge density wave alters the spatial symmetries in the crystal lattice of a related kagome superconductor.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 216-217 (2022)

doi: <https://doi.org/10.1038/d41586-022-00305-1>

References

1. Mielke, C. III *et al.* *Nature* **602**, 245–250 (2022).
2. Ortiz, B. R. *et al.* *Phys. Rev. Lett.* **125**, 247002 (2020).
3. Ortiz, B. R. *et al.* *Phys. Rev. Mater.* **5**, 034801 (2021).
4. Nie, L. *et al.* *Nature* <https://doi.org/10.1038/s41586-022-04493-8> (2022).
5. Moore, J. E. *Nature* **464**, 194–198 (2010).
6. Ortiz, B. R. *et al.* *Phys. Rev. Mater.* **3**, 094407 (2019).
7. Haldane, F. D. M. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
8. Varma, C. M. *Phys. Rev. B* **55**, 14554–14580 (1997).
9. Feng, X. Zhang, Y., Jiang, K. & Hu, J. *Phys. Rev. B* **104**, 165136 (2021).
10. Jiang, Y.-X. *et al.* *Nature Mater.* **20**, 1353–1357 (2021).
11. Ratcliff, N., Hallett, L., Ortiz, B. R., Wilson, S. D. & Harter, J. W. *Phys. Rev. Mater.* **5**, L111801 (2021).
12. Kang, M. *et al.* *Nature Phys.* <https://doi.org/10.1038/s41567-021-01451-5> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00305-1>

- NEWS AND VIEWS
- 08 February 2022

From the archive

Nature's pages feature the discovery of a key defence response, and a collection of items from an inventor highlights developments in photography.

100 years ago

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **602**, 217 (2022)

doi: <https://doi.org/10.1038/d41586-022-00306-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00306-0>

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

- NEWS AND VIEWS
- 10 January 2022

Inflammation in the gut is encoded by neurons in the brain

The nervous and immune systems interact in a bidirectional manner. It emerges that inflammation in the body activates brain cells that, when later reactivated, can trigger a recapitulation of the inflammatory response.

- [David Brea](#) ⁰ &
- [Henrique Veiga-Fernandes](#) ¹

Interactions between the nervous and the immune systems have been a topic of great interest over the past few decades. Neuronal signals can affect immune functions, and immune cells can modulate the activity of neurons in the brain and spinal cord, or in the rest of the body (known as the periphery), in health and disease^{1,2}. Writing in *Cell*, Koren *et al.*³ demonstrate that inflammation in the abdominal cavity results in the stimulation of certain neurons in a brain area called the insular cortex, or the insula. Artificial reactivation of these ‘immune-imprinted’ neurons is sufficient to generate organ-specific recall of inflammatory responses that resemble the initial inflammatory episode.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

Subscribe

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **602**, 217-218 (2022)

doi: <https://doi.org/10.1038/d41586-021-03802-x>

References

1. Chesne, J., Cardoso, V. & Veiga-Fernandes, H. *Mucosal Immunol.* **12**, 10–20 (2019).
2. Huh, J. R. & Veiga-Fernandes, H. *Nature Rev. Immunol.* **20**, 217–228 (2020).
3. Koren, T. *et al.* *Cell* **184**, 5902–5915 (2021).
4. Kipnis, J. *Science* **353**, 766–771 (2016).
5. Dantzer, R. *Physiol. Rev.* **98**, 477–504 (2018).
6. Cardoso, F. *et al.* *Nature* **597**, 410–414 (2021).
7. Godinho-Silva, C. *et al.* *Nature* **574**, 254–258 (2019).
8. Klose, C. S. N. & Veiga-Fernandes, H. *Eur. J. Immunol.* **51**, 1602–1614 (2021).
9. Veiga-Fernandes, H. & Mucida, D. *Cell* **165**, 801–811 (2016).
10. Wendeln, A. C. *et al.* *Nature* **556**, 332–338 (2018).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03802-x>

- Article
- [Published: 09 February 2022](#)

A white dwarf accreting planetary material determined from X-ray observations

- [Tim Cunningham](#) ORCID: orcid.org/0000-0001-7296-3533^{1,2},
- [Peter J. Wheatley](#) ORCID: orcid.org/0000-0003-1452-2240^{1,2},
- [Pier-Emmanuel Tremblay](#) ORCID: orcid.org/0000-0001-9873-0121^{1,2},
- [Boris T. Gänsicke](#) ORCID: orcid.org/0000-0002-2761-3005^{1,2},
- [George W. King](#) ORCID: orcid.org/0000-0002-3641-6636³,
- [Odette Toloza](#) ORCID: orcid.org/0000-0002-2398-719X^{4,5} &
- [Dimitri Veras](#) ORCID: orcid.org/0000-0001-8014-6162^{1,2}

Nature volume **602**, pages 219–222 (2022)

- 709 Accesses
- 527 Altmetric
- [Metrics details](#)

Subjects

- [Astrophysical disks](#)
- [Compact astrophysical objects](#)
- [Computational astrophysics](#)
- [High-energy astrophysics](#)
- [Stars](#)

Abstract

The atmospheres of a large proportion of white dwarf stars are polluted by heavy elements¹ that are expected to sink out of visible layers on short timescales^{2,3}. This has been interpreted as a signature of ongoing accretion of debris from asteroids⁴, comets⁵ and giant planets⁶. This scenario is supported by the detection of debris discs⁷ and transits of planetary fragments⁸ around some white dwarfs. However, photospheric metals are only indirect evidence for ongoing accretion, and the inferred accretion rates and parent body compositions heavily depend on models of diffusion and mixing processes within the white dwarf atmosphere^{9,10,11}. Here we report a 4.4σ detection of X-rays from a polluted white dwarf, G29–38. From the measured X-ray luminosity, we derive an instantaneous accretion rate of $\dot{M} = 1.63 \pm 0.40 \times 10^{-9} \text{ g s}^{-1}$, which is independent of stellar atmosphere models. This rate is higher than estimates from past studies of the photospheric abundances of G29–38, suggesting that convective overshoot may be needed to model the spectra of debris-accreting white dwarfs. We measure a low plasma temperature of $k_B T = 0.5 \pm 0.2 \text{ keV}$, corroborating the predicted bombardment solution for white dwarfs accreting at low accretion rates^{12,13}.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: X-ray and optical imaging of G29–38.



Fig. 2: Sky location of X-rays from G29–38.

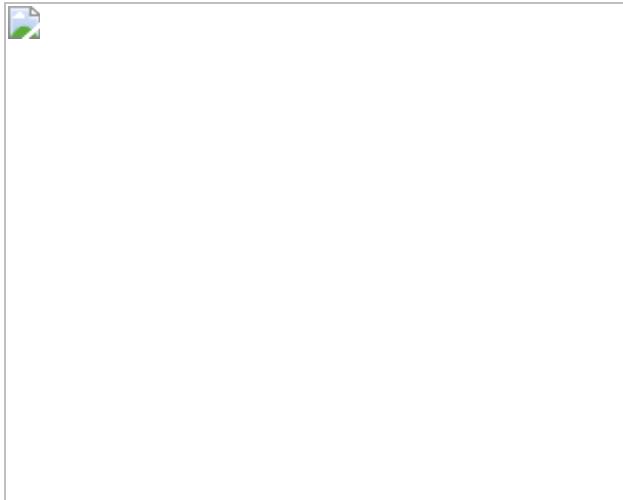
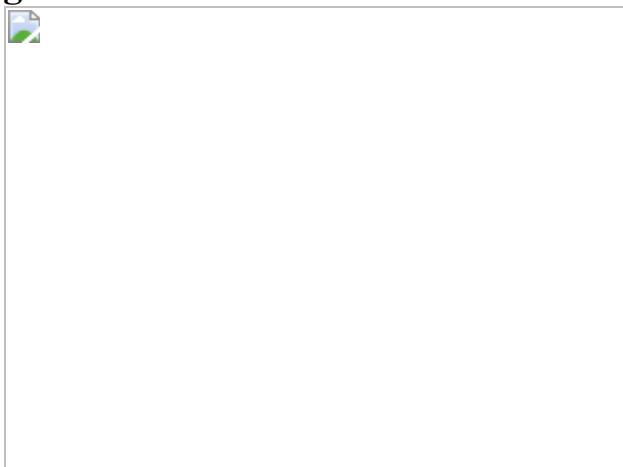


Fig. 3: Accretion rates inferred from the measured X-ray flux.



Data availability

The data that support the plots within this Article and other findings of this study are available from the *Chandra Data Archive*. The observation ID numbers are given in Extended Data Table [1](#).

Code availability

The official *Chandra* reduction software package CIAO, which includes merge_obs, wavdetect and XPSEC, is freely and publicly available (cxc.cfa.harvard.edu/ciao/), as is the Python package `scipy`.

References

1. Koester, D., Gänsicke, B. T. & Farihi, J. The frequency of planetary debris around young white dwarfs. *Astron. Astrophys.* **566**, A34 (2014).
2. Paquette, C., Pelletier, C., Fontaine, G. & Michaud, G. Diffusion coefficients for stellar plasmas. *Astrophys. J. Suppl.* **61**, 177–195 (1986).
3. Koester, D. Accretion and diffusion in white dwarfs. New diffusion timescales and applications to GD 362 and G 29-38. *Astron. Astrophys.* **498**, 517–525 (2009).
4. Zuckerman, B., Koester, D., Melis, C., Hansen, B. M. & Jura, M. The chemical composition of an extrasolar minor planet. *Astrophys. J.* **671**, 872–877 (2007).
5. Xu, S. et al. The chemical composition of an extrasolar Kuiper-Belt-Object. *Astrophys. J. Lett.* **836**, L7 (2017).
6. Gänsicke, B. T. et al. Accretion of a giant planet onto a white dwarf star. *Nature* **576**, 61–64 (2019).
7. Zuckerman, B. & Becklin, E. E. Excess infrared radiation from a white dwarf—an orbiting brown dwarf? *Nature* **330**, 138–140 (1987).
8. Vanderburg, A. et al. A disintegrating minor planet transiting a white dwarf. *Nature* **526**, 546–549 (2015).
9. Bauer, E. B. & Bildsten, L. Polluted white dwarfs: mixing regions and diffusion timescales. *Astrophys. J.* **872**, 96 (2019).
10. Cunningham, T., Tremblay, P.-E., Freytag, B., Ludwig, H.-G. & Koester, D. Convective overshoot and macroscopic diffusion in pure-hydrogen-atmosphere white dwarfs. *Mon. Not. R. Astron. Soc.* **488**, 2503–2522 (2019).
11. Heinonen, R. A. et al. Diffusion coefficients in the envelopes of white dwarfs. *Astrophys. J.* **896**, 2 (2020).

12. Kuijpers, J. & Pringle, J. E. Comments on radial white dwarf accretion. *Astron. Astrophys.* **114**, L4–L6 (1982).
13. Woelk, U. & Beuermann, K. Temperature structure of a particle-heated magnetic atmosphere. *Astron. Astrophys.* **280**, 169–173 (1993).
14. Graham, J. R., Matthews, K., Neugebauer, G. & Soifer, B. T. The infrared excess of G29–38: a brown dwarf or dust? *Astrophys. J.* **357**, 216 (1990).
15. Koester, D., Provencal, J. & Shipman, H. L. Metals in the variable DA G29-38. *Astron. Astrophys.* **320**, L57–L59 (1997).
16. Jura, M. A tidally disrupted asteroid around the white dwarf G29-38. *Astrophys. J. Lett.* **584**, L91–L94 (2003).
17. Xu, S., Jura, M., Koester, D., Klein, B. & Zuckerman, B. Elemental compositions of two extrasolar rocky planetesimals. *Astrophys. J.* **783**, 79 (2014).
18. Coutu, S. et al. Analysis of helium-rich white dwarfs polluted by heavy elements in the Gaia era. *Astrophys. J.* **885**, 74 (2019).
19. Mukai, K. X-ray emissions from accreting white dwarfs: a review. *Publ. Astron. Soc. Pacif.* **129**, 062001 (2017).
20. Patterson, J. & Raymond, J. C. X-ray emission from cataclysmic variables with accretion disks. I. Hard X-rays. *Astrophys. J.* **292**, 535–549 (1985).
21. Jura, M., Muno, M. P., Farihi, J. & Zuckerman, B. X-ray and infrared observations of two externally polluted white dwarfs. *Astrophys. J.* **699**, 1473–1479 (2009).
22. Farihi, J. et al. Magnetism, X-rays and accretion rates in WD 1145+017 and other polluted white dwarf systems. *Mon. Not. R. Astron. Soc.* **474**, 947–960 (2018).

23. Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P. & O'Dell, S. L. Chandra X-ray Observatory (CXO): overview. *Proc. SPIE* **4012**, 2–16 (2000).
24. Gaia Collaboration et al. Gaia Data Release 2. Summary of the contents and survey properties. *Astron. Astrophys.* **616**, A1 (2018).
25. Kylafis, N. D. & Lamb, D. Q. X-ray and UV radiation from accreting degenerate dwarfs. II. *Astrophys. J. Suppl.* **48**, 239–272 (1982).
26. Ghosh, P. & Lamb, F. K. Disk accretion by magnetic neutron stars. *Astrophys. J. Lett.* **223**, L83–L87 (1978).
27. Metzger, B. D., Rafikov, R. R. & Bochkarev, K. V. Global models of runaway accretion in white dwarf debris discs. *Mon. Not. R. Astron. Soc.* **423**, 505–528 (2012).
28. Farihi, J., Jura, M. & Zuckerman, B. Infrared signatures of disrupted minor planets at white dwarfs. *Astrophys. J.* **694**, 805–819 (2009).
29. Debes, J. H. & López-Morales, M. A second look at the metal line variability of G29-38. *Astrophys. J. Lett.* **677**, L43 (2008).
30. *The Chandra Proposers' Observatory Guide* (Chandra X-ray Center, accessed 30 October 2021);
<https://cxc.cfa.harvard.edu/proposer/POG/html/index.html>
31. Fruscione, A. et al. CIAO: Chandra's data analysis system. *Proc. SPIE* **6270**, 62701V (2006).
32. Li, T. P. & Ma, Y. Q. Analysis methods for results in gamma-ray astronomy. *Astrophys. J.* **272**, 317–324 (1983).
33. Gaia Collaboration et al. Gaia Early Data Release 3. Summary of the contents and survey properties. *Astron. Astrophys.* **649**, A1 (2021).
34. Kraft, R. P., Burrows, D. N. & Nousek, J. A. Determination of confidence limits for experiments with low numbers of counts. *Astrophys. J.* **374**, 344–355 (1991).

35. Freeman, P. E., Kashyap, V., Rosner, R. & Lamb, D. Q. A wavelet-based algorithm for the spatial analysis of Poisson data. *Astrophys. J. Suppl.* **138**, 185–218 (2002).
36. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
37. Arnaud, K. A. XSPEC: the first ten years. *Astron. Soc. Pac. Conf. Ser.* **101**, 17–20 (1996).
38. Foster, A. R., Ji, L., Smith, R. K. & Brickhouse, N. S. Updated atomic data and calculations for x-ray spectroscopy. *Astrophys. J.* **756**, 128 (2012).
39. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. the chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
40. McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
41. Redfield, S. & Linsky, J. L. The three-dimensional structure of the warm local interstellar medium. II. The Colorado model of the local interstellar cloud. *Astrophys. J.* **534**, 825–837 (2000).
42. Cash, W. Parameter estimation in astronomy through application of the likelihood ratio. *Astrophys. J.* **228**, 939–947 (1979).
43. Patterson, J., Zuckerman, B., Becklin, E. E., Tholen, D. J. & Hawarden, T. the infrared and optical pulsations of G29-38. *Astrophys. J.* **374**, 330–339 (1991).
44. Kleinman, S. J. et al. Understanding the cool DA white dwarf pulsator, G29-38. *Astrophys. J.* **495**, 424–434 (1998).
45. McCleery, J. et al. Gaia white dwarfs within 40 pc II: the volume-limited Northern hemisphere sample. *Mon. Not. R. Astron. Soc.* **499**, 1890–1908 (2020).

46. Gentile Fusillo, N. P. et al. A catalogue of white dwarfs in Gaia EDR3. Preprint at <https://arxiv.org/abs/2106.07669> (2021).
47. Gianninas, A., Bergeron, P. & Ruiz, M. T. A spectroscopic survey and analysis of bright, hydrogen-rich white dwarfs. *Astrophys. J.* **743**, 138 (2011).
48. Tremblay, P. E. & Bergeron, P. Spectroscopic analysis of DA white dwarfs: Stark broadening of hydrogen lines including nonideal effects. *Astrophys. J.* **696**, 1755–1770 (2009).
49. Tremblay, P.-E., Ludwig, H.-G., Steffen, M. & Freytag, B. Spectroscopic analysis of DA white dwarfs with 3D model atmospheres. *Astron. Astrophys.* **559**, A104 (2013).
50. Koester, D., Kepler, S. O. & Irwin, A. W. New white dwarf envelope models and diffusion. Application to DQ white dwarfs. *Astron. Astrophys.* **635**, A103 (2020).
51. Freytag, B., Ludwig, H.-G. & Steffen, M. Hydrodynamical models of stellar convection. The role of overshoot in DA white dwarfs, A-type stars, and the Sun. *Astron. Astrophys.* **313**, 497–516 (1996).
52. Kupka, F., Zaussinger, F. & Montgomery, M. H. Mixing and overshooting in surface convection zones of DA white dwarfs: first results from ANTARES. *Mon. Not. R. Astron. Soc.* **474**, 4660–4671 (2018).

Acknowledgements

This research was supported by a Leverhulme Trust Grant (ID RPG-2020-366). P.J.W., B.T.G. and P.-E.T. were supported by the UK STFC consolidated grant no. ST/T000406/1. P.-E.T. received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme no. 677706 (WD3D). B.T.G. was supported by a Leverhulme Research Fellowship, O.T. was supported by a Leverhulme Trust Research Project Grant and FONDECYT project no.

32103, and D.V. acknowledges the support of the STFC via an Ernest Rutherford Fellowship (grant no. ST/P003850/1). This research has made use of data obtained from the *Chandra Data Archive* and the *Chandra Source Catalog*, and software provided by the Chandra X-ray Center (CXC) in the application packages CIAO and Sherpa. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the arvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant no. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant no. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory and the Gordon and Betty Moore Foundation.

Author information

Affiliations

1. Department of Physics, The University of Warwick, Coventry, UK

Tim Cunningham, Peter J. Wheatley, Pier-Emmanuel Tremblay, Boris T. Gänsicke & Dimitri Veras

2. Centre for Exoplanets and Habitability, University of Warwick,
Coventry, UK

Tim Cunningham, Peter J. Wheatley, Pier-Emmanuel Tremblay, Boris
T. Gänsicke & Dimitri Veras

3. Department of Astronomy, University of Michigan, Ann Arbor, MI,
USA

George W. King

4. Departamento de Física, Universidad Técnica Federico Santa María,
Valparaíso, Chile

Odette Toloza

5. Millennium Nucleus for Planet Formation (NPF), Valparaíso, Chile

Odette Toloza

Contributions

T.C. performed most of the data analysis and led the writing of the manuscript. P.J.W. contributed to the original observing proposal, the data analysis and the writing of the manuscript. P.-E.T and B.T.G. contributed to the writing of the manuscript and discussions on the various constraints on accretion rates. G.W.K., O.T. and D.V. contributed to the original observing proposal and to discussions throughout the project.

Corresponding author

Correspondence to [Tim Cunningham](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Andrew Vanderburg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 Estimate of sky density from a source detection algorithm.](#)

Output of the wavdetect source detection algorithm using the recommended significance threshold, sigthresh, of $s_1 = 10^{-6}$. **a–c**, are the results for the science bands used in this study; soft, soft + medium and broad, respectively, with 1, 2, and 3 counts shown in cyan, magenta, and white, respectively. **a–c**, Full field of view of the S3 chip on ACIS-S with sources identified by wavdetect shown in green. **d, e**, A magnified view of the vicinity near the target for the soft and soft + medium bands respectively, where the white square has sides of length 30 arcsec and is centered on the target coordinates. The source at the sky location of G29–38 is detected in the soft and soft + medium band images. The number of sources and corresponding sky density for each band can be found in Extended Data Table 3. **f**, Broadband (0.5–7.0 keV) events recorded near the sky position of the only *Gaia* source (EDR3 source ID: 2661110815469507072) in the full CCD image which has a detected X-ray source in the broadband image within 5 arcsec. The output of wavdetect gives the coordinates for the center of the detected source in (RA, DEC) as $(352.182358 \pm 0.000022,$

$+5.341529 \pm 0.000020$), shown in the figure with a cyan cross. The coordinates of the *Gaia* EDR3 source, corrected for proper motion, are $(352.18248533 \pm 0.00000007, +5.34150200 \pm 0.00000008)$, indicated in the figure with a white cross. The *Chandra* source is offset compared to the expected *Gaia* position by $(-0.46 \pm 0.08, 0.10 \pm 0.07)$, which is consistent with the offset of our target from the expected position (see Fig. 2).

Extended Data Fig. 2 Monte Carlo aperture photometry.

a, c, The blue points show the positions of the approximately 100,000 test apertures, each 1 arcsec in radius, used to sample 100 arcsec around the target. The absolute number of test apertures, after removing those that fell within a masked region, is shown in the panels. The sky coordinates of all recorded events that fall within a masked region are shown in orange. **b, d**, The normalized histogram shows the fraction of test apertures with event counts equal to or greater than that of a given bin. The Monte Carlo was performed on the soft (**a, b**) and soft + medium (**c, d**) bands. The soft band analysis has 0.001% of test apertures returning four counts, enabling us to rule out chance alignment at 4.4σ .

Extended Data Fig. 3 Spectral modelling of observed X-ray events.

a, Bottom, in units of instrumental counts we show the five recorded events (black) and six best-fit spectral models assuming Solar³⁹ (magenta), bulk Earth⁴⁰ (blue) and photospheric¹⁷ (red) abundances, with either the vvapec isothermal (solid) and mkcfow cooling flow (dotted) plasma models. We also indicate the dominant metal emission lines (O, Mg, Si and Fe) from the isothermal, photospheric abundance model. Middle, in real flux units, we show the synthetic spectra for the photospheric abundances with the isothermal (solid) and cooling flow (dotted) plasma models. The modelling suggests the most likely origin of the source photon at 1.3 keV was a Mg transition. Top, the effective area of the ACIS-S detector is shown in green, hatch. The absence of harder X-ray events (>2.0 keV) in the *Chandra* observations demonstrates that the plasma emission spectrum is very soft. **b**, Spectral energy distribution of the best-fit isothermal (blue) and cooling

flow (orange) plasma models with bulk Earth abundances down to the extreme ultraviolet (EUV) energy regime. Also shown are the standard *Chandra* science bands; soft, medium and hard. Both models provide a convergent fit within the *Chandra* ACIS-S passband, but the cooling flow provides a more physical and larger estimate of the lower-energy flux. **c**, The X-ray flux measured in 3 bands: 0.5–7.0 keV, 0.3–7.0 keV and 0.0136–100 keV, using the cooling flow model for the photospheric abundances¹⁷ is shown in open diamonds. The filled horizontal bands show the 68% and 90% confidence intervals on the X-ray accretion rate, which is computed using Equation (1). The X-ray accretion rates are computed using Equation (1), with $A = 1$, $R_{\text{WD}} = 0.0129R_{\odot}$, and $M_{\text{WD}} = 0.6M_{\odot}$. The spectroscopic accretion rates (solid lines) are the same as shown in Fig. 3, with the 1σ uncertainty shown with dashed lines. Also shown in solid circles (blue and orange) are the previously published inferred accretion rates for G29–38, based on photospheric abundances from spectroscopic observations^{17,28}.

Extended Data Fig. 4 Limit on cyclotron emission cooling as source of additional luminosity.

a, An estimate for the total luminosity from the measured X-ray luminosity, accounting for cyclotron emission cooling. We compare the measured plasma temperature, $k_{\text{B}}T_{\text{X}} = (0.61 \pm 0.28)$ keV, given by the cooling flow model and photospheric abundances, with the critical plasma temperature, T_B , above which cyclotron emission cooling dominates, defined by equation 10 from ref. ²². The authors provide the ratio $\langle T \rangle_{\text{X}} / \langle T \rangle_B \approx L_{\text{tot}} / L_X$ as an estimate of, for a range of accretion rates and global magnetic field strengths, the predicted increase in total luminosity compared to X-ray luminosity if the plasma temperature is sufficient to be dominated by cyclotron emission cooling. The horizontal dotted line indicates $T_X/T_B = 1$, where no correction is expected below this. The vertical dotted line indicates the 3σ upper limit on the magnetic field strength from FORS2 spectropolarimetric observations²². The solid lines indicate the increase in total luminosity when compared to the observed X-ray luminosity. **b**, Predicted additional luminosity for an assumed global magnetic field at the 3σ limit (1.5 kG) across the full range of plasma temperatures and accretion

rates calculated in this work (see Extended Data Tables 5 & 6). White space indicates no additional luminosity. The upper plasma temperature from the cooling flow model and accretion rate derived from the isothermal plasma model is shown (solid) along with the 68% uncertainty (dashed). Even at the observational upper limit, the predicted increase due to cyclotron emission cooling is a factor of $\{1.1\}_{-0.1}^{+0.6}$.

Extended Data Table 1 *Chandra* observations of G29–38

Extended Data Table 2 Statistical significance of source detection

Extended Data Table 3 Sky density of sources from the wavdetect source detection algorithm

Extended Data Table 4 Confidence interval on *Chandra* ACIS-S count rate

Extended Data Table 5 Plasma temperature from spectral modelling using XSPEC

Extended Data Table 6 Best-fit X-ray properties of G29–38

Supplementary information

[**Peer Review File**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Cunningham, T., Wheatley, P.J., Tremblay, PE. *et al.* A white dwarf accreting planetary material determined from X-ray observations. *Nature* **602**, 219–222 (2022). <https://doi.org/10.1038/s41586-021-04300-w>

- Received: 29 July 2021
- Accepted: 01 December 2021

- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04300-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/s41586-021-04300-w>

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

- Article
- [Published: 09 February 2022](#)

Outracing champion Gran Turismo drivers with deep reinforcement learning

- [Peter R. Wurman](#) ORCID: orcid.org/0000-0001-9349-0624¹,
- [Samuel Barrett](#)¹,
- [Kenta Kawamoto](#)²,
- [James MacGlashan](#)¹,
- [Kaushik Subramanian](#)³,
- [Thomas J. Walsh](#)¹,
- [Roberto Capobianco](#) ORCID: orcid.org/0000-0002-2219-215X³,
- [Alisa Devlic](#)³,
- [Franziska Eckert](#)³,
- [Florian Fuchs](#) ORCID: orcid.org/0000-0002-1072-3718³,
- [Leilani Gilpin](#) ORCID: orcid.org/0000-0002-9741-2014¹,
- [Piyush Khandelwal](#)¹,
- [Varun Kompella](#)¹,
- [HaoChih Lin](#)³,
- [Patrick MacAlpine](#)¹,
- [Declan Oller](#)¹,
- [Takuma Seno](#)²,
- [Craig Sherstan](#)¹,
- [Michael D. Thomure](#)¹,
- [Houmehr Aghabozorgi](#)¹,
- [Leon Barrett](#) ORCID: orcid.org/0000-0001-9896-3259¹,
- [Rory Douglas](#)¹,
- [Dion Whitehead](#)¹,

- [Peter Dürr](#) ORCID: [orcid.org/0000-0002-3840-5009³](https://orcid.org/0000-0002-3840-5009),
- [Peter Stone¹](#),
- [Michael Spranger](#) ORCID: [orcid.org/0000-0001-9443-7008²](https://orcid.org/0000-0001-9443-7008) &
- [Hiroaki Kitano²](#)

Nature volume **602**, pages 223–228 (2022)

- 6491 Accesses
- 999 Altmetric
- [Metrics details](#)

Subjects

- [Applied mathematics](#)
- [Computer science](#)

Abstract

Many potential applications of artificial intelligence involve making real-time decisions in physical systems while interacting with humans. Automobile racing represents an extreme example of these conditions; drivers must execute complex tactical manoeuvres to pass or block opponents while operating their vehicles at their traction limits¹. Racing simulations, such as the PlayStation game Gran Turismo, faithfully reproduce the non-linear control challenges of real race cars while also encapsulating the complex multi-agent interactions. Here we describe how we trained agents for Gran Turismo that can compete with the world's best e-sports drivers. We combine state-of-the-art, model-free, deep reinforcement learning algorithms with mixed-scenario training to learn an integrated control policy that combines exceptional speed with impressive tactics. In addition, we construct a reward function that enables the agent to be competitive while adhering to racing's important, but under-specified, sportsmanship rules. We demonstrate the capabilities of our agent, Gran Turismo Sophy, by winning a head-to-head competition against four of the world's best Gran Turismo drivers. By describing how we trained

championship-level racers, we demonstrate the possibilities and challenges of using these techniques to control complex dynamical systems in domains where agents must respect imprecisely defined human norms.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Training.



Fig. 2: Ablations.

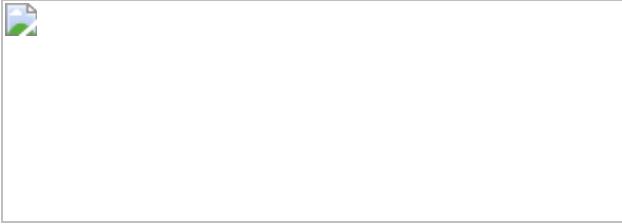


Fig. 3: Results.



Data availability

There are no static data associated with this project. All data are generated from scratch by the agent each time it learns. Videos of the races are available at https://sonyai.github.io/gt_sophy_public.

Code availability

Pseudocode detailing the training process and algorithms used is available as a supplement to this article. The agent interface in GT is not enabled in commercial versions of the game; however, Polyphony Digital has provided a small number of universities and research facilities outside Sony access to the API and is considering working with other groups.

References

1. Milliken, W. F. et al. *Race Car Vehicle Dynamics* Vol. 400 (Society of Automotive Engineers, 1995).
2. Mnih, V. et al. Playing Atari with deep reinforcement learning. Preprint at <https://arxiv.org/abs/1312.5602> (2013).
3. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
4. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
5. Vinyals, O. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* **575**, 350–354 (2019).
6. Berner, C. et al. Dota 2 with large scale deep reinforcement learning. Preprint at <https://arxiv.org/abs/1912.06680> (2019).
7. Laurence, V. A., Goh, J. Y. & Gerdes, J. C. In *2017 American Control Conference (ACC)* 5586–5591 (IEEE, 2017).
8. Spielberg, N. A., Brown, M., Kapania, N. R., Kegelman, J. C. & Gerdes, J. C. Neural network vehicle models for high-performance automated driving. *Sci. Robot.* **4**, eaaw1975 (2019).
9. Burke, K. Data makes it beta: Roborace returns for second season with updateable self-driving vehicles powered by NVIDIA DRIVE. *The*

Official NVIDIA Blog

<https://blogs.nvidia.com/blog/2020/10/29/robورace-second-season-nvidia-drive/> (2020).

10. Leporati, G. No driver? no problem—this is the Indy Autonomous Challenge. *Ars Technica* <https://arstechnica.com/cars/2021/07/a-science-fair-or-the-future-of-racing-the-indy-autonomous-challenge/> (2021).
11. Williams, G., Drews, P., Goldfain, B., Rehg, J. M. & Theodorou, E. A. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* 1433–1440 (IEEE, 2016).
12. Williams, G., Drews, P., Goldfain, B., Rehg, J. M. & Theodorou, E. A. Information-theoretic model predictive control: theory and applications to autonomous driving. *IEEE Trans. Robot.* **34**, 1603–1622 (2018).
13. Pan, Y. et al. In *Proc. Robotics: Science and Systems XIV* (eds Kress-Gazit, H., Srinivasa, S., Howard, T. & Atanasov, N.) <https://doi.org/10.15607/RSS.2018.XIV.056> (Carnegie Mellon Univ., 2018).
14. Pan, Y. et al. Imitation learning for agile autonomous driving. *Int. J. Robot. Res.* **39**, 286–302 (2020).
15. Amazon Web Services. AWS DeepRacer League. <https://aws.amazon.com/deepracer/league/> (2019).
16. Pyeatt, L. D. & Howe, A. E. Learning to race: experiments with a simulated race car. In *Proc. Eleventh International FLAIRS Conference* 357–361 (AAAI, 1998).
17. Chaperot, B. & Fyfe, C. In *2006 IEEE Symposium on Computational Intelligence and Games* 181–186 (IEEE, 2006).
18. Cardamone, L., Loiacono, D. & Lanzi, P. L. In *Proc. 11th Annual Conference on Genetic and Evolutionary Computation* 1179–1186

(ACM, 2009).

19. Cardamone, L., Loiacono, D. & Lanzi, P. L. In *2009 IEEE Congress on Evolutionary Computation* 2622–2629 (IEEE, 2009).
20. Loiacono, D., Prete, A., Lanzi, L. & Cardamone, L. In *IEEE Congress on Evolutionary Computation* 1–8 (IEEE, 2010).
21. Jaritz, M., de Charette, R., Toromanoff, M., Perot, E. & Nashashibi, F. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* 2070–2075 (IEEE, 2018).
22. Weiss, T. & Behl, M. In *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)* 1163–1168 (IEEE, 2020).
23. Weiss, T., Babu, V. S. & Behl, M. In *NeurIPS 2020 Workshop on Machine Learning for Autonomous Driving* (NeurIPS, 2020).
24. Fuchs, F., Song, Y., Kaufmann, E., Scaramuzza, D. & Dürr, P. Super-human performance in Gran Turismo Sport using deep reinforcement learning. *IEEE Robot. Autom. Lett.* **6**, 4257–4264 (2021).
25. Song, Y., Lin, H., Kaufmann, E., Dürr, P. & Scaramuzza, D. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2021).
26. Theodosis, P. A. & Gerdes, J. C. In *Dynamic Systems and Control Conference* Vol. 45295, 235–241 (American Society of Mechanical Engineers, 2012).
27. Funke, J. et al. In *2012 IEEE Intelligent Vehicles Symposium* 541–547 (IEEE, 2012).
28. Kritayakirana, K. & Gerdes, J. C. Autonomous vehicle control at the limits of handling. *Int. J. Veh. Auton. Syst.* **10**, 271–296 (2012).
29. Bonkowski, J. Here's what you missed from the Indy Autonomous Challenge main event. *Autoweek*

<https://www.autoweek.com/racing/more-racing/a38069263/what-missed-indy-autonomous-challenge-main-event/> (2021).

30. Rutherford, S. J. & Cole, D. J. Modelling nonlinear vehicle dynamics with neural networks. *Int. J. Veh. Des.* **53**, 260–287 (2010).
31. Pomerleau, D. A. In *Robot Learning* (eds Connell, J. H. & Mahadevan, S.) 19–43 (Springer, 1993).
32. Togelius, J. & Lucas, S. M. In *2006 IEEE International Conference on Evolutionary Computation* 1187–1194 (IEEE, 2006).
33. Schwarting, W. et al. Deep latent competition: learning to race using visual control policies in latent space. Preprint at <https://arxiv.org/abs/2102.09812> (2021).
34. Gozli, D. G., Bavelier, D. & Pratt, J. The effect of action video game playing on sensorimotor learning: evidence from a movement tracking task. *Hum. Mov. Sci.* **38**, 152–162 (2014).
35. Davids, K., Williams, A. M. & Williams, J. G. *Visual Perception and Action in Sport* (Routledge, 2005).
36. Haarnoja, T., Zhou, A., Abbeel, P. & Levine, S. In *Proc. 35th International Conference on Machine Learning* 1856–1865 (PMLR, 2018).
37. Haarnoja, T. et al. Soft actor-critic algorithms and applications. Preprint at <https://arxiv.org/abs/1812.05905> (2018).
38. Mnih, V. et al. In *Proc. 33rd International Conference on Machine Learning* 1928–1937 (PMLR, 2016).
39. Dabney, W., Rowland, M., Bellemare, M. G. & Munos, R. In *32nd AAAI Conference on Artificial Intelligence* (AAAI, 2018).
40. Lin, L.-J. *Reinforcement Learning for Robots Using Neural Networks*. Dissertation, Carnegie Mellon Univ. (1993).

41. Siu, H. C. et al. Evaluation of human-AI teams for learned and rule-based agents in Hanabi. Preprint at <https://arxiv.org/abs/2107.07630> (2021).
42. Tesauro, G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.* **6**, 215–219 (1994).
43. Devore, J. L. *Probability and Statistics for Engineering and the Sciences* 6th edn (Brooks/Cole, 2004).
44. Xia, L., Zhou, Z., Yang, J. & Zhao, Q. DSAC: distributional soft actor critic for risk-sensitive reinforcement learning. Preprint at <https://arxiv.org/abs/2004.14547> (2020).
45. Fujimoto, S., van Hoof, H. & Meger, D. In *Proc. 35th International Conference on Machine Learning* 1587–1596 (PMLR, 2018).
46. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
47. Liu, Z., Li, X., Kang, B. & Darrell, T. In *International Conference on Learning Representations* (ICLR, 2021).
48. Kingma, D. P. & Ba, J. In *International Conference on Learning Representations* (ICLR, 2015).
49. Cassirer, A. et al. Reverb: a framework for experience replay. Preprint at <https://arxiv.org/abs/2102.04736> (2021).
50. Narvekar, S., Sinapov, J., Leonetti, M. & Stone, P. In *Proc. 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)* (2016).

Acknowledgements

We thank K. Yamauchi, S. Takano, A. Hayashi, C. Ferreira, N. Nozawa, T. Teramoto, M. Hakim, K. Yamada, S. Sakamoto, T. Ueda, A. Yago, J.

Nakata and H. Imanishi at Polyphony Digital for making the Gran Turismo franchise, providing support throughout the project and organizing the Race Together events on 2 July 2021 and 21 October 2021. We also thank U. Gallizzi, J. Beltran, G. Albowicz, R. Abdul-ahad and the staff at CGEI for access to their PlayStation Now network to train agents and their help building the infrastructure for our experiments. We benefited from the advice of T. Grossenbacher, a retired competitive GT driver. Finally, we thank E. Kato Marcus and E. Ohshima of Sony AI, who managed the partnership activities with Polyphony Digital and Sony Interactive Entertainment.

Author information

Affiliations

1. Sony AI, New York, NY, USA

Peter R. Wurman, Samuel Barrett, James MacGlashan, Thomas J. Walsh, Leilani Gilpin, Piyush Khandelwal, Varun Kompella, Patrick MacAlpine, Declan Oller, Craig Sherstan, Michael D. Thomure, Houmehr Aghabozorgi, Leon Barrett, Rory Douglas, Dion Whitehead & Peter Stone

2. Sony AI, Tokyo, Japan

Kenta Kawamoto, Takuma Seno, Michael Spranger & Hiroaki Kitano

3. Sony AI, Zürich, Switzerland

Kaushik Subramanian, Roberto Capobianco, Alisa Devlic, Franziska Eckert, Florian Fuchs, HaoChih Lin & Peter Dürr

Contributions

P.R.W. managed the project. S.B., K.K., P.K., J.M., K.S. and T.J.W. led the research and development efforts. R.C., A.D., F.E., F.F., L.G., V.K., H.L., P.M., D.O., C.S., T.S. and M.D.T. participated in the research and the

development of GT Sophy and the AI libraries. H.A., L.B., R.D. and D.W. built the research platform that connected to CGEI's PlayStation network. P.S. provided executive support and technical and research advice and P.D. provided executive support and technical advice. H.K. and M.S. conceived and set up the project, provided executive support, resources and technical advice and managed stakeholders.

Corresponding author

Correspondence to [Peter R. Wurman](#).

Ethics declarations

Competing interests

P.R.W. and other team members have submitted US provisional patent application 63/267,136 covering aspects of the scenario training techniques described in this paper.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

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 Seaside and Sarthe training.](#)

Kudos Prime data from global time-trial challenges on Seaside (**a** and **b**) and Sarthe (**c** and **d**), with the cars used in the competition. Note that these histograms represent the single best lap time for more than 12,000 individual players on Seaside and almost 9,000 on Sarthe. In both cases, the secondary diagrams compare the top five human times to a histogram of 100 laps by the 2 July 2021 time-trial version of GT Sophy. In both cases, the data show that GT Sophy was reliably superhuman, with all 100 laps better than the best human laps. Not surprisingly, it takes longer for the agent to train on the much longer Sarthe course, taking 48 h to reach the 99th percentile of human performance. **e**, Histogram of a snapshot of the ERB during training on Sarthe on the basis of the scenario breakdown in Fig. [1f](#). The *x* axis is the course position and the stacked colours represent the number of samples that were collected in that region from each scenario. In a more condensed format than Fig. [1f](#), **f** and **g** show the sections of Seaside and Maggiore that were used for skill training.

Extended Data Fig. 2 Time trial on Sarthe.

An analysis of Igor Fraga’s best lap in the time-trial test compared with GT Sophy’s lap. **a**, Areas of the track where Igor lost time with respect to GT Sophy. Corner 20, highlighted in yellow, shows an interesting effect common to the other corners in that Igor seems to catch up a little by braking later, but then loses time because he has to brake longer and comes out of the corner slower. Igor’s steering controls (**b**) and Igor’s throttle and braking (**c**) compared with GT Sophy on corner 20. Through the steering wheel and brake pedal, Igor is able to give smooth, 60-Hz signals compared with GT Sophy’s 10-Hz action rate.

Extended Data Fig. 3 Policy selection.

An illustration of the process by which policies were selected to run in the final race. Starting on the left side of the diagram, thousands of policies were generated and saved during the experiments. They were first filtered in the experiment to select the subset on the Pareto frontier of a simple evaluation criteria trading off lap time versus off-course and collision metrics. The selected policies were run through a series of tests evaluating their overall racing performance against a common set of opponents and

their performance on a variety of hand-crafted skill tests. The results were ranked and human judgement was applied to select a small number of candidate policies. These policies were matched up in round-robin, policy-versus-policy competitions. The results were again analysed by the human committee for overall team scores and collision metrics. The best candidate policies were run in short races against test drivers at Polyphony Digital. Their subjective evaluations were included in the final decisions on which policies to run in the October 2021 event.

Extended Data Table 1 Reward weights

Supplementary information

Supplementary Information

This file contains more details about the training procedures and algorithms in the form of pseudocode. It also contains several tables that detail the hyperparameters used in training.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Wurman, P.R., Barrett, S., Kawamoto, K. *et al.* Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature* **602**, 223–228 (2022). <https://doi.org/10.1038/s41586-021-04357-7>

- Received: 09 August 2021
- Accepted: 15 December 2021

- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04357-7>

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

[**Neural networks overtake humans in *Gran Turismo* racing game**](#)

- J. Christian Gerdes

News & Views 09 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04357-7>

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

- Article
- [Published: 09 February 2022](#)

Evidence for the association of triatomic molecules in ultracold $^{23}\text{Na}^{40}\text{K} + ^{40}\text{K}$ mixtures

- [Huan Yang](#) [ORCID: orcid.org/0000-0001-9160-8385](#)^{1,2,3} na1,
- [Xin-Yao Wang](#)^{1,2,3,4,5} na1,
- [Zhen Su](#)^{1,2,3},
- [Jin Cao](#)^{1,2,3},
- [De-Chao Zhang](#)^{1,2,3},
- [Jun Rui](#)^{1,2,3},
- [Bo Zhao](#) [ORCID: orcid.org/0000-0002-4950-1880](#)^{1,2,3},
- [Chun-Li Bai](#) [ORCID: orcid.org/0000-0002-5253-3656](#)^{4,5} &
- [Jian-Wei Pan](#) [ORCID: orcid.org/0000-0002-6100-5142](#)^{1,2,3}

[Nature](#) volume **602**, pages 229–233 (2022)

- 1245 Accesses
- 44 Altmetric
- [Metrics details](#)

Subjects

- [Atomic and molecular interactions with photons](#)
- [Chemical physics](#)
- [Ultracold gases](#)

Abstract

Ultracold assembly of diatomic molecules has enabled great advances in controlled chemistry, ultracold chemical physics and quantum simulation with molecules^{1,2,3}. Extending the ultracold association to triatomic molecules will offer many new research opportunities and challenges in these fields. A possible approach is to form triatomic molecules in a mixture of ultracold atoms and diatomic molecules by using a Feshbach resonance between them^{4,5}. Although ultracold atom–diatomic-molecule Feshbach resonances have been observed recently^{6,7}, using these resonances to form triatomic molecules remains challenging. Here we report on evidence of the association of triatomic molecules near the Feshbach resonance between $^{23}\text{Na}^{40}\text{K}$ molecules in the rovibrational ground state and ^{40}K atoms. We apply a radio-frequency pulse to drive the free-bound transition in ultracold mixtures of $^{23}\text{Na}^{40}\text{K}$ and ^{40}K and monitor the loss of $^{23}\text{Na}^{40}\text{K}$ molecules. The association of triatomic molecules manifests itself as an additional loss feature in the radio-frequency spectra, which can be distinguished from the atomic loss feature. The observation that the distance between the association feature and the atomic transition changes with the magnetic field provides strong evidence for the formation of triatomic molecules. The binding energy of the triatomic molecules is estimated from the measurements. Our work contributes to the understanding of the complex ultracold atom–molecule Feshbach resonances and may open up an avenue towards the preparation and control of ultracold triatomic molecules.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Illustration of the rf association of triatomic molecules in the vicinity of a Feshbach resonance between $^{23}\text{Na}^{40}\text{K}$ and ^{40}K .

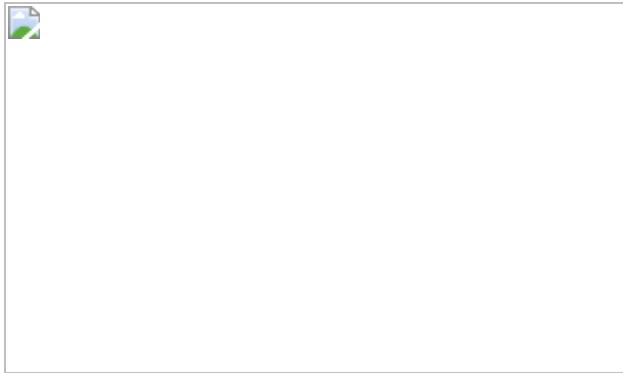


Fig. 2: Rf spectra measured at different magnetic fields near the Feshbach resonance between $|0, 0, -3/2, -3\rangle$ and $|9/2, -7/2\rangle$.

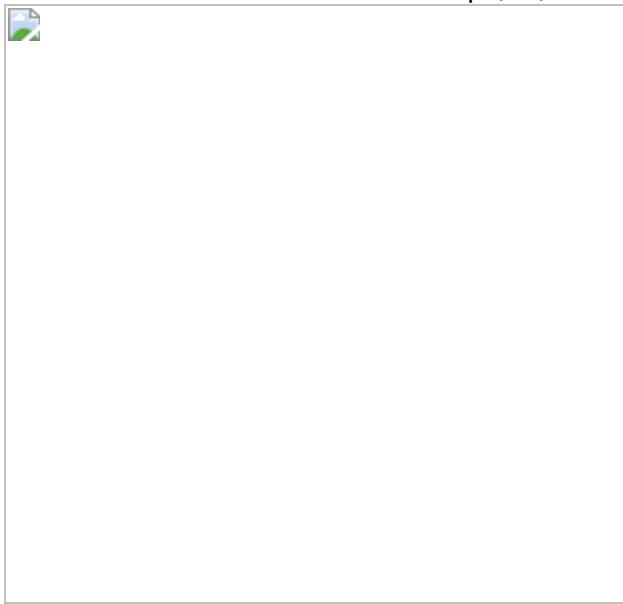


Fig. 3: Binding energies of the triatomic molecules as a function of the magnetic field.



Data availability

All data generated or analysed during this study are included in this published article (and its supplementary information files). [Source data](#) are provided with this paper.

References

1. Krems, R. V. Cold controlled chemistry. *Phys. Chem. Chem. Phys.* **10**, 4079–4092 (2008).
2. Carr, L. D., DeMille, D., Krems, R. V. & Ye, J. Cold and ultracold molecules: science, technology and applications. *New J. Phys.* **11**, 055049 (2009).
3. Quéméner, G. & Julienne, P. S. Ultracold molecules under control. *Chem. Rev.* **112**, 4949–5011 (2012).
4. Chin, C., Grimm, R., Julienne, P. & Tiesinga, E. Feshbach resonances in ultracold gases. *Rev. Mod. Phys.* **82**, 1225–1286 (2010).
5. Köhler, T., Góral, K. & Julienne, P. S. Production of cold molecules via magnetically tunable Feshbach resonances. *Rev. Mod. Phys.* **78**, 1311–1361 (2006).
6. Yang, H. et al. Observation of magnetically tunable Feshbach resonances in ultracold $^{23}\text{Na}^{40}\text{K}$ + ^{40}K collisions. *Science* **363**, 261–264 (2019).
7. Wang, X.-Y. et al. Magnetic Feshbach resonances in collisions of $^{23}\text{Na}^{40}\text{K}$ with ^{40}K . *New J. Phys.* **23**, 115010 (2021).
8. Shuman, E. S., Barry, J. F. & DeMille, D. Laser cooling of a diatomic molecule. *Nature* **467**, 820–823 (2010).
9. Barry, J. F. et al. Magneto-optical trapping of a diatomic molecule. *Nature* **512**, 286–289 (2014).
10. Cheuk, L. W. et al. Λ -enhanced imaging of molecules in an optical trap. *Phys. Rev. Lett.* **121**, 083201 (2018).

11. Caldwell, L. et al. Deep laser cooling and efficient magnetic compression of molecules. *Phys. Rev. Lett.* **123**, 033202 (2019).
12. Ding, S., Wu, Y., Finneran, I. A., Burau, J. J. & Ye, J. Sub-Doppler cooling and compressed trapping of YO molecules at μK temperatures. *Phys. Rev. X* **10**, 021049 (2020).
13. Ni, K.-K. et al. A high phase-space-density gas of polar molecules. *Science* **322**, 231–235 (2008).
14. Molony, P. K. et al. Creation of ultracold $^{87}\text{Rb}^{133}\text{Cs}$ molecules in the rovibrational ground state. *Phys. Rev. Lett.* **113**, 255301 (2014).
15. Takekoshi, T. et al. Ultracold dense samples of dipolar RbCs molecules in the rovibrational and hyperfine ground state. *Phys. Rev. Lett.* **113**, 205301 (2014).
16. Park, J. W., Will, S. A. & Zwierlein, M. W. Ultracold dipolar gas of fermionic $^{23}\text{Na}^{40}\text{K}$ molecules in their absolute ground state. *Phys. Rev. Lett.* **114**, 205302 (2015).
17. Guo, M. et al. Creation of an ultracold gas of ground-state dipolar $^{23}\text{Na}^{87}\text{Rb}$ molecules. *Phys. Rev. Lett.* **116**, 205303 (2016).
18. Rvachov, T. M. et al. Long-lived ultracold molecules with electric and magnetic dipole moments. *Phys. Rev. Lett.* **119**, 143001 (2017).
19. Seeßelberg, F. et al. Modeling the adiabatic creation of ultracold polar $^{23}\text{Na}^{40}\text{K}$ molecules. *Phys. Rev. A* **97**, 013405 (2018).
20. Voges, K. K. et al. Ultracold gas of bosonic $^{23}\text{Na}^{39}\text{K}$ ground-state molecules. *Phys. Rev. Lett.* **125**, 083401 (2020).
21. Liu, L. R. et al. Building one molecule from a reservoir of two atoms. *Science* **360**, 900–903 (2018).
22. He, X. et al. Coherently forming a single molecule in an optical trap. *Science* **370**, 331–335 (2020).

23. Ospelkaus, S. et al. Quantum-state controlled chemical reactions of ultracold potassium-rubidium molecules. *Science* **327**, 853–857 (2010).
24. Hu, M.-G. et al. Direct observation of bimolecular reactions of ultracold KRb molecules. *Science* **366**, 1111–1115 (2019).
25. Son, H., Park, J. J., Ketterle, W. & Jamison, A. O. Collisional cooling of ultracold molecules. *Nature* **580**, 197–200 (2020).
26. DeMarco, L. et al. A degenerate Fermi gas of polar molecules. *Science* **363**, 853–856 (2019).
27. Valtolina, G. et al. Dipolar evaporation of reactive molecules to below the Fermi temperature. *Nature* **588**, 239–243 (2020).
28. Kozyryev, I. et al. Sisyphus laser cooling of a polyatomic molecule. *Phys. Rev. Lett.* **118**, 173201 (2017).
29. Zeppenfeld, M. et al. Sisyphus cooling of electrically trapped polyatomic molecules. *Nature* **491**, 570–573 (2012).
30. Mitra, D. et al. Direct laser cooling of a symmetric top molecule. *Science* **369**, 1366–1369 (2020).
31. Qiu, M. et al. Observation of Feshbach resonances in the $\text{F} + \text{H}_2 \rightarrow \text{HF} + \text{H}$ reaction. *Science* **311**, 1440–1443 (2006).
32. Vogels, S. N. et al. Imaging resonances in low-energy NO-He inelastic collisions. *Science* **350**, 787–790 (2015).
33. Chefdeville, S. et al. Observation of partial wave resonances in low-energy O_2-H_2 inelastic collisions. *Science* **341**, 1094–1096 (2013).
34. Klein, A. et al. Directly probing anisotropy in atom–molecule collisions through quantum scattering resonances. *Nat. Phys.* **13**, 35–38 (2017).

35. Regal, C. A., Ticknor, C., Bohn, J. L. & Jin, D. S. Creation of ultracold molecules from a Fermi gas of atoms. *Nature* **424**, 47–50 (2003).
36. Herbig, J. et al. Preparation of a pure molecular quantum gas. *Science* **301**, 1510–1513 (2003).
37. Thompson, S. T., Hodby, E. & Wieman, C. E. Ultracold molecule production via a resonant oscillating magnetic field. *Phys. Rev. Lett.* **95**, 190404 (2005).
38. Ospelkaus, C. et al. Ultracold heteronuclear molecules in a 3D optical lattice. *Phys. Rev. Lett.* **97**, 120402 (2006).
39. Lompe, T. et al. Radio-frequency association of Efimov trimers. *Science* **330**, 940–944 (2010).
40. Nakajima, S., Horikoshi, M., Mukaiyama, T., Naidon, P. & Ueda, M. Measurement of an Efimov trimer binding energy in a three-component mixture of ${}^6\text{Li}$. *Phys. Rev. Lett.* **106**, 143201 (2011).
41. Machtey, O., Shotan, Z., Gross, N. & Khaykovich, L. Association of Efimov trimers from a three-atom continuum. *Phys. Rev. Lett.* **108**, 210406 (2012).
42. Christianen, A., Zwierlein, M. W., Groenenboom, G. C. & Karman, T. Photoinduced two-body loss of ultracold molecules. *Phys. Rev. Lett.* **123**, 123402 (2019).
43. Gregory, P. D., Blackmore, J. A., Bromley, S. L. & Cornish, S. L. Loss of ultracold ${}^{87}\text{Rb}{}^{133}\text{Cs}$ molecules via optical excitation of long-lived two-body collision complexes. *Phys. Rev. Lett.* **124**, 163402 (2020).
44. Liu, Y. et al. Photo-excitation of long-lived transient intermediates in ultracold reactions. *Nat. Phys.* **16**, 1132–1136 (2020).
45. Nichols, M. A. et al. Detection of long-lived complexes in ultracold atom-molecule collisions. Preprint at <https://arxiv.org/abs/2105.14960> (2021).

46. Balakrishnan, N. On the role of van der Waals interaction in chemical reactions at low temperatures. *J. Chem. Phys.* **121**, 5563–5566 (2004).
47. Krems, R. & Dalgarno, A. Quantum-mechanical theory of atom-molecule and molecular collisions in a magnetic field: spin depolarization. *J. Chem. Phys.* **120**, 2296–2307 (2004).
48. Lepers, M. & Dulieu, O. Long-range interactions between ultracold atoms and molecules including atomic spin–orbit. *Phys. Chem. Chem. Phys.* **13**, 19106–19113 (2011).
49. Żuchowski, P. S., Kosicki, M., Kodrycka, M. & Soldán, P. van der Waals coefficients for systems with ultracold polar alkali-metal molecules. *Phys. Rev. A* **87**, 022706 (2013).
50. Gao, B. Angular-momentum-insensitive quantum-defect theory for diatomic systems. *Phys. Rev. A* **64**, 010701 (2001).
51. Gao, B., Tiesinga, E., Williams, C. J. & Julienne, P. S. Multichannel quantum-defect theory for slow atomic collisions. *Phys. Rev. A* **72**, 042719 (2005).
52. Crubellier, A. & Luc-Koenig, E. Threshold effects in the photoassociation of cold atoms: R^{-6} model in the Milne formalism. *J. Phys. B At. Mol. Opt. Phys.* **39**, 1417 (2006).
53. Crubellier, A., González-Férez, R., Koch, C. P. & Luc-Koenig, E. Asymptotic model for shape resonance control of diatomics by intense non-resonant light. *New J. Phys.* **17**, 045020 (2015).
54. Gao, B. Multichannel quantum-defect theory for anisotropic interactions. Preprint at <https://arxiv.org/abs/2008.08018> (2020).
55. Son, H. et al. Control of reactive collisions by quantum interference. Preprint at <https://arxiv.org/abs/2109.03944> (2021).
56. González-Martinez, M. L. & Hutson, J. M. Effect of hyperfine interactions on ultracold molecular collisions: $\text{NH}(^3\Sigma^-)$ with $\text{Mg}(^1\text{s})$ in

- magnetic fields. *Phys. Rev. A* **84**, 052706 (2011).
57. Morita, M., Kłos, J. & Tscherbul, T. V. Full-dimensional quantum scattering calculations on ultracold atom-molecule collisions in magnetic fields: the role of molecular vibrations. *Phys. Rev. Res.* **2**, 043294 (2020).
58. Hauser, A. W. & Ernst, W. E. Homo- and heteronuclear alkali metal trimers formed on helium nanodroplets. Part I. Vibronic spectra simulations based on ab initio calculations. *Phys. Chem. Chem. Phys.* **13**, 18762–18768 (2011).
59. Schnabel, J., Kampschulte, T., Rupp, S., Hecker Denschlag, J. & Köhn, A. Towards photoassociation processes of ultracold rubidium trimers. *Phys. Rev. A* **103**, 022820 (2021).
60. Pérez-Ros, J., Lepers, M. & Dulieu, O. Theory of long-range ultracold atom-molecule photoassociation. *Phys. Rev. Lett.* **115**, 073201 (2015).
61. Park, J. W. et al. Quantum degenerate Bose-Fermi mixture of chemically different atomic species with widely tunable interactions. *Phys. Rev. A* **85**, 051602 (2012).
62. Liu, L. et al. Observation of interference between resonant and detuned STIRAP in the adiabatic creation of $^{23}\text{Na}^{40}\text{K}$ molecules. *Phys. Rev. Lett.* **122**, 253201 (2019).

Acknowledgements

We thank P. Julienne, J. Hutson, J. Ye, B. Gao and J. Doyle for helpful discussions. This work was supported by the National Key R&D Program of China (under grant number 2018YFA0306502), the National Natural Science Foundation of China (under grant numbers 11521063 and 11904355), the Chinese Academy of Sciences, the Anhui Initiative in Quantum Information Technologies, the Shanghai Municipal Science and Technology Major Project (grant number 2019SHZDZX01) and the Shanghai Rising-Star Program (grant number 20QA1410000).

Author information

Author notes

1. These authors contributed equally: Huan Yang, Xin-Yao Wang

Affiliations

1. Hefei National Laboratory for Physical Sciences at the Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui, China

Huan Yang, Xin-Yao Wang, Zhen Su, Jin Cao, De-Chao Zhang, Jun Rui, Bo Zhao & Jian-Wei Pan

2. Shanghai Branch, CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai, China

Huan Yang, Xin-Yao Wang, Zhen Su, Jin Cao, De-Chao Zhang, Jun Rui, Bo Zhao & Jian-Wei Pan

3. Shanghai Research Center for Quantum Sciences, Shanghai, China

Huan Yang, Xin-Yao Wang, Zhen Su, Jin Cao, De-Chao Zhang, Jun Rui, Bo Zhao & Jian-Wei Pan

4. Beijing National Laboratory for Molecular Sciences, Key Laboratory of Molecular Nanostructure and Nanotechnology, CAS Research/Education Center for Excellence in Molecular Sciences, Institute of Chemistry, Chinese Academy of Sciences, Beijing, China

Xin-Yao Wang & Chun-Li Bai

5. University of Chinese Academy of Sciences, Beijing, China

Xin-Yao Wang & Chun-Li Bai

Contributions

B.Z., C.-L.B. and J.-W.P. conceived the experiments. H.Y., X.-Y.W., Z.S., J.C., D.-C.Z. and J.R. carried out the experiments. All authors analysed the data and contributed to the writing of the paper. B.Z., C.-L.B. and J.-W.P. supervised the work.

Corresponding authors

Correspondence to [Bo Zhao](#), [Chun-Li Bai](#) or [Jian-Wei Pan](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

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 The magnetic field as a function of the hold time.](#)

After preparation of the $^{23}\text{Na}^{40}\text{K}$ molecules, the magnetic field is then ramped to the target value B_t in 3 ms, after which we wait 18 ms for the magnetic field to stabilize. The rf association pulse is applied between 21 ms and 51 ms. The magnetic field is measured by rf spectroscopy. The dashed lines represent the target magnetic field B_t .

[Source data](#)

Extended Data Fig. 2 Comparison of the rf spectra in the continuous-wave dipole trap and in the intensity-modulated dipole trap.

a, b, The rf spectra measured in the continuous-wave dipole trap. We use two overlapping Gaussian functions to fit the data (red solid line), where the Gaussian function for the atomic loss feature is centred at the atomic transition. Each point represents the average of 6–10 measurements and error bars represent the standard error of the mean. As a comparison, the rf spectra measured in the intensity-modulated optical dipole trap are shown in **c** and **d**. It can be seen that the association feature is better resolved in the modulated dipole trap.

[Source data](#)

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Yang, H., Wang, XY., Su, Z. *et al.* Evidence for the association of triatomic molecules in ultracold $^{23}\text{Na}^{40}\text{K} + ^{40}\text{K}$ mixtures. *Nature* **602**, 229–233 (2022). <https://doi.org/10.1038/s41586-021-04297-2>

- Received: 07 May 2021
- Accepted: 29 November 2021
- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04297-2>

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/s41586-021-04297-2>

- Article
- [Published: 09 February 2022](#)

Real-space visualization of intrinsic magnetic fields of an antiferromagnet

- [Yuji Kohno](#)¹,
- [Takehito Seki](#)^{2,3},
- [Scott D. Findlay](#) [ORCID: orcid.org/0000-0003-4862-4827](#)⁴,
- [Yuichi Ikuhara](#) [ORCID: orcid.org/0000-0003-3886-005X](#)^{2,5} &
- [Naoya Shibata](#) [ORCID: orcid.org/0000-0003-3548-5952](#)^{2,5,6}

[Nature](#) volume **602**, pages 234–239 (2022)

- 1724 Accesses
- 28 Altmetric
- [Metrics details](#)

Subjects

- [Magnetic properties and materials](#)
- [Transmission electron microscopy](#)

Abstract

Characterizing magnetic structures down to atomic dimensions is central to the design and control of nanoscale magnetism in materials and devices. However, real-space visualization of magnetic fields at such dimensions has

been extremely challenging. In recent years, atomic-resolution differential phase contrast scanning transmission electron microscopy (DPC STEM)¹ has enabled direct imaging of electric field distribution even inside single atoms². Here we show real-space visualization of magnetic field distribution inside antiferromagnetic haematite ($\alpha\text{-Fe}_2\text{O}_3$) using atomic-resolution DPC STEM in a magnetic-field-free environment³. After removing the phase-shift component due to atomic electric fields and improving the signal-to-noise ratio by unit-cell averaging, real-space visualization of the intrinsic magnetic fields in $\alpha\text{-Fe}_2\text{O}_3$ is realized. These results open a new possibility for real-space characterization of many magnetic structures.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Schematic illustration of $\alpha\text{-Fe}_2\text{O}_3$ structure and atomic-resolution ADF STEM image.

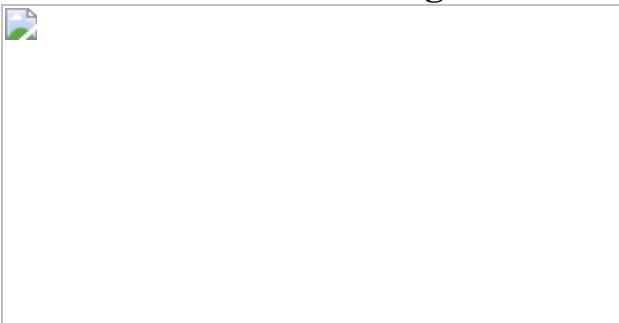


Fig. 2: Statistical image analysis of the high-S/N B -field-filtered DPC image.

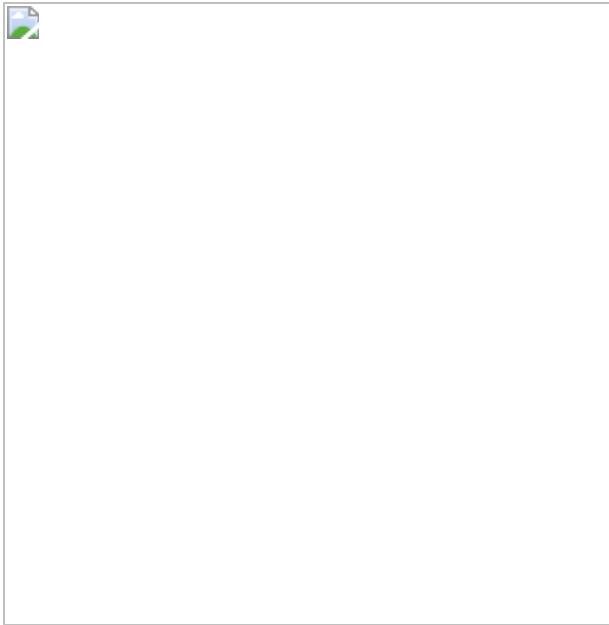


Fig. 3: Unit-cell-averaged and tiled ADF image, projected magnetic field vector colour map and reconstructed phase image of $\alpha\text{-Fe}_2\text{O}_3$ observed along the $\bar{[1\bar{1}20]}$ direction.

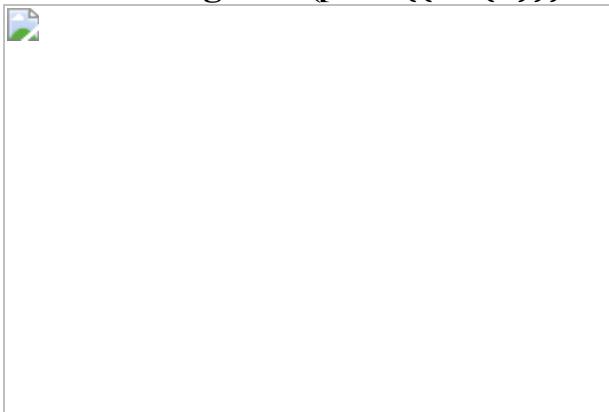
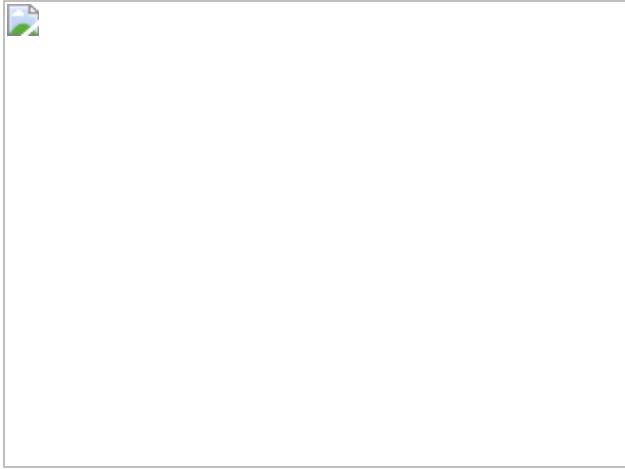


Fig. 4: Unit-cell-averaged and tiled ADF image, projected magnetic field vector colour map and reconstructed phase image of $\alpha\text{-Fe}_2\text{O}_3$.



Data availability

The data supporting the findings of this study are available within the paper and its [Supplementary Information](#). The raw DPC images after 25-image averaging but before kernel filtering are available at
<https://github.com/sigma-users/kernel-filter>.

Code availability

The custom-designed code used for kernel image filtering is available at
<https://github.com/sigma-users/kernel-filter>.

References

1. Shibata, N. et al. Differential phase-contrast microscopy at atomic resolution. *Nat. Phys.* **8**, 611–615 (2012).
2. Shibata, N. et al. Electric field imaging of single atoms. *Nat. Commun.* **8**, 15631 (2017).
3. Shibata, N. et al. Atomic resolution electron microscopy in a magnetic field free environment. *Nat. Commun.* **10**, 2308 (2019).
4. Aharonov, Y. & Bohm, D. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491 (1959).

5. Zweck, J. Imaging of magnetic and electric fields by electron microscopy. *J. Phys. Condens. Matter* **28**, 403001 (2016).
6. Chapman, J. N. The investigation of magnetic domain structures in thin foils by electron microscopy. *J. Phys. D* **17**, 623–647 (1984).
7. Matsumoto, T. et al. Direct observation of $\Sigma 7$ grain boundary core structure in magnetic skyrmion lattice. *Sci. Adv.* **2**, e1501280 (2016).
8. Chen, C. L. et al. Direct determination of atomic structure and magnetic coupling of magnetite twin boundaries. *ACS Nano* **12**, 2662–2668 (2018).
9. Shibata, N. et al. New area detector for atomic-resolution scanning transmission electron microscopy. *J. Electron Microsc.* **59**, 473–479 (2010).
10. Sánchez-Santolino, G. et al. Probing the internal atomic charge density distributions in real space. *ACS Nano* **12**, 8875–8881 (2018).
11. Shindo, D. & Murakami, Y. Electron holography of magnetic materials. *J. Phys. D* **41**, 183002 (2008).
12. Rother, A. & Scheerschmidt, K. Relativistic effects in elastic scattering of electrons in TEM. *Ultramicroscopy* **109**, 154–160 (2009).
13. Edström, A., Lubk, A. & Rusz, J. Quantum mechanical treatment of atomic-resolution differential phase contrast imaging of magnetic materials. *Phys. Rev. B* **99**, 174428 (2019).
14. Sawada, H. et al. Correction of higher order geometrical aberration by triple 3-fold astigmatism field. *J. Electron Microsc.* **58**, 341–347 (2009).
15. Morin, F. J. Magnetic susceptibility of $\alpha\text{-Fe}_2\text{O}_3$ and $\alpha\text{-Fe}_2\text{O}_3$ with added titanium. *Phys. Rev.* **78**, 819–820 (1950).

16. Shull, C. G., Strauser, W. A. & Wollan, E. O. Neutron diffraction by paramagnetic and antiferromagnetic substances. *Phys. Rev.* **83**, 333–345 (1951).
17. Brok, E. et al. Spin orientation in solid solution hematite-ilmenite. *Am. Mineral.* **102**, 1234–1243 (2017).
18. Ishizuka, A. et al. Boundary-artifact-free determination of potential distribution from differential phase contrast signals. *Microscopy* **66**, 397–405 (2017).
19. Egerton, R. F. *Electron Energy-Loss Spectroscopy in the Electron Microscope* 3rd edn (Springer, 2011).
20. Tanigaki, T. et al. Magnetic field observations in CoFeB/Ta layers with 0.67-nm resolution by electron holography. *Sci. Rep.* **7**, 16598 (2017).
21. Close, R., Chen, Z., Shibata, N. & Findlay, S. D. Towards quantitative, atomic-resolution reconstruction of the electrostatic potential via differential phase contrast using electrons. *Ultramicroscopy* **159**, 124–137 (2015).
22. Jones, L. et al. Smart Align—a new tool for robust non-rigid registration of scanning microscope data. *Adv. Struct. Chem. Imaging* **1**, 8 (2015).
23. Iakoubovskii, K., Mitsuishi, K., Nakayama, Y. & Furuya, K. Mean free path of inelastic electron scattering in elemental solids and oxides using transmission electron microscopy: atomic number dependent oscillatory behavior. *Phys. Rev. B* **77**, 104102 (2008).
24. Loudon, J. C. Antiferromagnetism in NiO observed by transmission electron diffraction. *Phys. Rev. Lett.* **109**, 267204 (2012).
25. Brown, P. J. In *International Tables for Crystallography Vol. C* (ed. Prince, E.) 454–461 (Kluwer Academic, 2006).

26. Krén, E., Szabó, P. & Konczos, G. Neutron diffraction studies on $(1-x)\text{Fe}_2\text{O}_3-x\text{Rh}_2\text{O}_3$ system. *Phys. Lett.* **19**, 103–104 (1965).
27. Müller, K. et al. Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction. *Nat. Commun.* **5**, 5653 (2014).
28. Lubk, A. & Zweck, J. Differential phase contrast: an integral perspective. *Phys. Rev. A* **91**, 023805 (2015).
29. Lugg, N. R., Neish, M. J., Findlay, S. D. & Allen, L. J. Practical aspects of removing the effects of elastic and thermal diffuse scattering from spectroscopic data for single crystals. *Microsc. Microanal.* **20**, 1078–1089 (2014).
30. Campanini, M., Nasi, L., Albertini, F. & Erni, R. Disentangling nanoscale electric and magnetic fields by time-reversal operation in differential phase-contrast STEM. *Appl. Phys. Lett.* **117**, 154102 (2020).

Acknowledgements

This work was supported by the JST SENTAN grant number JPMJSN14A1, Japan. A part of this work was supported by the JSPS KAKENHI grant numbers 20H05659, 19H05788 and 17H06094. A part of this work was supported by the Research Hub for Advanced Nano Characterization, The University of Tokyo, under the support of ‘Nanotechnology Platform’ by MEXT, Japan (grant number JPMXP09A21UT0259). We also thank the Corporate Sponsored Research Program ‘Next generation electron microscopy’, School of Engineering, The University of Tokyo. T.S. acknowledges support from JST-PRESTO (JPMJPR21AA), JSPS KAKENHI grant number 20K15014, and the Kazato Research Foundation. This research was partly supported under the Discovery Projects funding scheme of the Australian Research Council (project numbers DP160102338 and FT190100619).

Author information

Affiliations

1. JEOL Ltd., Tokyo, Japan

Yuji Kohno

2. Institute of Engineering Innovation, School of Engineering, The University of Tokyo, Tokyo, Japan

Takehito Seki, Yuichi Ikuhara & Naoya Shibata

3. PRESTO, Japan Science and Technology Agency, Kawaguchi, Japan

Takehito Seki

4. School of Physics and Astronomy, Monash University, Melbourne, Victoria, Australia

Scott D. Findlay

5. Nanostructures Research Laboratory, Japan Fine Ceramic Center, Nagoya, Japan

Yuichi Ikuhara & Naoya Shibata

6. Quantum-Phase Electronics Center (QPEC), The University of Tokyo, Tokyo, Japan

Naoya Shibata

Contributions

Y.K., T.S. and N.S. designed the study and wrote the paper. Y.K. performed the STEM experiments and image analysis. T.S. performed image simulation, theoretical analysis and prepared the TEM specimen. S.D.F. and Y.I. contributed to the discussion and comments. N.S. directed the study.

Corresponding author

Correspondence to [Naoya Shibata](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Kernel filter for subtracting the electric field signal component.](#)

Schematic illustration of the kernel filter. In this filter, DPC signals of four nearby points (denoted 2, 4, 6 and 8) are averaged and then subtracted from that of the centre point (denoted 0). This operation is illustrated in the right schematic. The numbers in the schematic show the weighting factors at each point in this kernel filter. The distance between the points corresponds to the distance between the Fe atomic dumbbells. This filter subtracts the electric field component and reinforces the magnetic field component at the centre Fe atomic columns.

Extended Data Fig. 2 Comparison between experimental DPC(x, y) images obtained using different kernel filters.

a, Kernel filter (raw) for obtaining the same DPC images as the original DPC images. **b**, Kernel filter (E -field) for subtracting the magnetic field component. Therefore, the obtained DPC images should correspond to the electric field images. It is seen that the original DPC image in **a** and the electric field image in **b** appear to be identical. This is because the main signal component of the original DPC is the atomic electric field. **c**, Kernel filter (B -field) for subtracting the electric field component as already shown in Extended Data Fig. 1. **d**, Kernel filter (null) for subtracting both the electric and magnetic field components. In the field vector colour map, the field vectors are plotted for the beam-deflection angle range 0–0.6 mrad in the raw and E -field kernel filtered images, whereas those in the B -field and null kernel filtered images are plotted for the range 0–50 μ rads. In the DPC(x, y) images, the grey scale beam-deflection angle ranges are shown on the colour bars. In the fast Fourier transforms (FFTs), the scale bar corresponds to $(0.4 \text{ nm})^{-1}$. The intensity scales are the same in all the FFTs. The inset in the FFT of DPC(y) shows the weak double periodicity spots related to the antiferromagnetic structure.

Extended Data Fig. 3 The standard-error maps for raw, E -field-filtered and B -field-filtered DPC(x, y) images (after binning down to 128×128 px).

Here, the propagation of uncertainty from the raw DPC images are fully considered for the kernel-filtered images.

Extended Data Fig. 4 The average/standard error maps for B -field-filtered DPC(x, y) images.

After binning down to 128×128 px.

Extended Data Fig. 5 Nearby-unit-averaged B -field-filtered DPC images.

The kernels used to average nearby units are shown schematically in the right panels. The centre panel shows the corresponding simulated B -filtered DPC images including the effect of finite electron dose used in the experiment. In the field vector colour maps (left), the field vectors are plotted with beam-deflection angle ranges shown by the inset (top right). Note especially that the structural regularity that emerges in the experimental images is similar across regions further apart than the width of the averaging region. Scale bars are 1 nm.

Extended Data Fig. 6 Statistical image analysis of the high-S/N kernel-filtered DPC(x, y) images.

Plot of the averages and the standard errors of the A-site and B-site averaged beam-deflection angles for B -field, E -field and null-filtered high-S/N DPC images. The standard errors are shown as the error bars. The inset shows a magnified portion of the plot around (0, 0).

Extended Data Fig. 7 Amplitudes of Fourier components of the unit-cell-averaged DPC images.

Only the Fourier components corresponding to the double periodicity due to the magnetic structure are plotted. The amplitudes that are non-zero with statistical significance (the error bars are shown by $\pm 2SE$) are plotted in blue, whereas the remainder are plotted in red. The statistically significant non-zero amplitudes of Fourier components with the highest spatial frequency can be found at $6.83 \text{ nm}^{-1} = 1/(1.46 \text{ \AA})$, indicated by the vertical arrow. Thus, the unit-cell-averaged magnetic field reconstruction contains statistically reliable spatial frequency information up to around $(1.46 \text{ \AA})^{-1}$.

Extended Data Fig. 8 Spin orientations.

a–c, Red arrows show the $\langle [\bar{1}\bar{1}20] \rangle$ -projected spin orientations assuming that they have the tilt angle θ_0 to the (0001) plane and perpendicular to the $\langle [\bar{1}\bar{1}20] \rangle$ (**a**), $\langle [\bar{1}2\bar{1}0] \rangle$ (**b**) and $\langle [2\bar{1}\bar{1}0] \rangle$ (**c**) directions. The orientation parameters used in equation (7) become $\sin\tau = 1$. In **a**, $\eta = \theta_0$; in **b** and **c**, $\langle \sin \tau \rangle$

$=\sqrt{\{\sin \}^2\{\theta _0\}+\frac{1}{4}\{\cos \}^2\{\theta _0\}}$) and $\eta = \arctan(2\tan\theta_0)$. In the present $([\bar{1}\bar{1}20])$ projection, **b** and **c** cannot be distinguished.

Extended Data Fig. 9 Simulated magnetic phase-shift images (infinite dose case) by systematically changing the out-of-plane component (in angle) of the spin direction from the (0001) basal plane.

a–e, The out-of-plane components are 0° (**a**), 10° (**b**), 20° (**c**), 30° (**d**) and 90° (**e**), for the antiferromagnetic domain shown in Extended Data Fig. 8b. It is seen that the magnetic phase-shift images sensitively change depending on the out-of-plane components of the spin direction. **f–j**, The out-of-plane components are 0° (**f**), 10° (**g**), 20° (**h**), 30° (**i**) and 90° (**j**), for the antiferromagnetic domain shown in Extended Data Fig. 8a. The best match with the experimental phase-shift image shown in Fig. 3 is **c**.

Extended Data Fig. 10 Comparison between simulated magnetic DPC image approximations and a simulated magnetic DPC image using the multislice method.

a, Simulated, approximate magnetic field vector colour map, in which a purely magnetic potential is placed at each Fe atomic position and averaged over a simulated probe profile as it would evolve in the absence of the sample (that is, without scattering). **b**, Simulated, approximate magnetic field vector colour map, in which a purely magnetic potential is placed at each Fe atomic position and averaged over the simulated probe profile resulting from assuming scattering from the electrostatic potential only. **c**, Residual electric field component calculated by the first term in equation (12). **d**, Simulated, approximate magnetic DPC image formed by adding **c** to the magnetic DPC image shown in **b**; that is, as per both terms in equation (12). **e**, Magnetic field vector colour map obtained by applying B -field kernel filter to the multislice image simulation (infinite dose version of Fig. 3e). All these images assume infinite dose imaging condition.

Supplementary information

Supplementary Information

This file contains Supplementary Text and Supplementary Figs. 1–22.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kohno, Y., Seki, T., Findlay, S.D. *et al.* Real-space visualization of intrinsic magnetic fields of an antiferromagnet. *Nature* **602**, 234–239 (2022).
<https://doi.org/10.1038/s41586-021-04254-z>

- Received: 17 July 2020
- Accepted: 16 November 2021
- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04254-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/s41586-021-04254-z>

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

- Article
- [Published: 09 February 2022](#)

Ferroelectric incommensurate spin crystals

- [Dorin Rusu](#)¹,
- [Jonathan J. P. Peters](#)^{1,2},
- [Thomas P. A. Hase](#) [ORCID: orcid.org/0000-0001-5274-5942](#)¹,
- [James A. Gott](#) [ORCID: orcid.org/0000-0001-8521-5104](#)¹,
- [Gareth A. A. Nisbet](#) [ORCID: orcid.org/0000-0003-2314-5862](#)³,
- [Jörg Strempfer](#) [ORCID: orcid.org/0000-0002-6443-0283](#)⁴,
- [Daniel Haskel](#)⁴,
- [Samuel D. Seddon](#) [ORCID: orcid.org/0000-0001-8900-9308](#)¹,
- [Richard Beanland](#) [ORCID: orcid.org/0000-0003-1749-4134](#)¹,
- [Ana M. Sanchez](#) [ORCID: orcid.org/0000-0002-8230-6059](#)¹ &
- [Marin Alexe](#) [ORCID: orcid.org/0000-0002-0386-3026](#)¹

[Nature](#) volume **602**, pages 240–244 (2022)

- 1753 Accesses
- 100 Altmetric
- [Metrics details](#)

Subjects

- [Ferroelectrics and multiferroics](#)
- [Surfaces, interfaces and thin films](#)

Abstract

Ferroics, especially ferromagnets, can form complex topological spin structures such as vortices¹ and skyrmions^{2,3} when subjected to particular electrical and mechanical boundary conditions. Simple vortex-like, electric-dipole-based topological structures have been observed in dedicated ferroelectric systems, especially ferroelectric–insulator superlattices such as $\text{PbTiO}_3/\text{SrTiO}_3$, which was later shown to be a model system owing to its high depolarizing field^{4,5,6,7,8}. To date, the electric dipole equivalent of ordered magnetic spin lattices driven by the Dzyaloshinskii–Moriya interaction (DMI)^{9,10} has not been experimentally observed. Here we examine a domain structure in a single PbTiO_3 epitaxial layer sandwiched between SrRuO_3 electrodes. We observe periodic clockwise and anticlockwise ferroelectric vortices that are modulated by a second ordering along their toroidal core. The resulting topology, supported by calculations, is a labyrinth-like pattern with two orthogonal periodic modulations that form an incommensurate polar crystal that provides a ferroelectric analogue to the recently discovered incommensurate spin crystals in ferromagnetic materials^{11,12,13}. These findings further blur the border between emergent ferromagnetic and ferroelectric topologies, clearing the way for experimental realization of further electric counterparts of magnetic DMI-driven phases.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Unit-cell-scale electric dipole topologies.

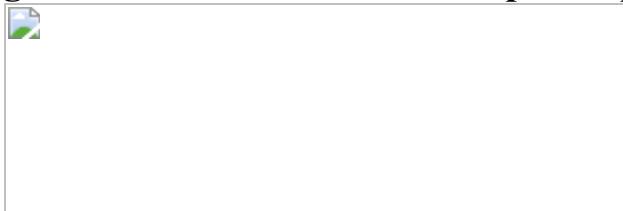


Fig. 2: Macroscopic ordering.

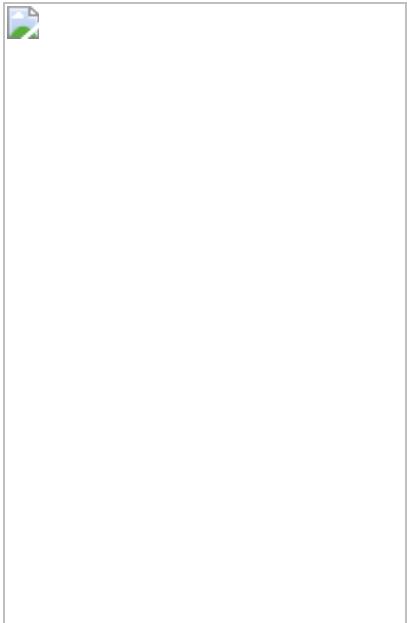


Fig. 3: Plan-view TEM imaging.

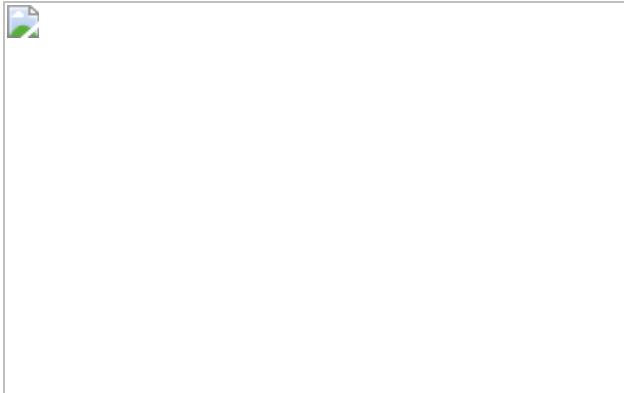
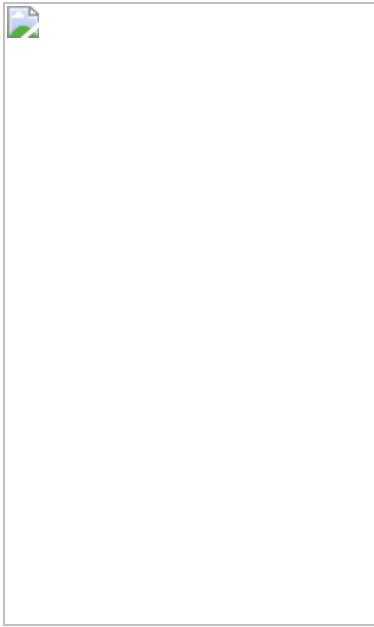


Fig. 4: Double modulation.



Data availability

The data that support the findings of this study are available at the University of Warwick open access research repository (<http://wrap.warwick.ac.uk/159776/>) or from the corresponding author on reasonable request.

References

1. Wachowiak, A. et al. Direct observation of internal spin structure of magnetic vortex cores. *Science* **298**, 577–580 (2002).
2. Seki, S., Yu, X. Z., Ishiwata, S. & Tokura, Y. Observation of skyrmions in a multiferroic material. *Science* **336**, 198–201 (2012).
3. Yu, X. Z. et al. Real-space observation of a two-dimensional skyrmion crystal. *Nature* **465**, 901–904 (2010).
4. Yadav, A. K. et al. Observation of polar vortices in oxide superlattices. *Nature* **530**, 198–201 (2016).

5. Hsu, S.-L. et al. Emergence of the vortex state in confined ferroelectric heterostructures. *Adv. Mater.* **31**, 1901014–1901022 (2019).
6. Shafer, P. et al. Emergent chirality in the electric polarization texture of titanate superlattices. *Proc. Natl Acad. Sci. USA* **115**, 915–920 (2018).
7. Gruverman, A. et al. Vortex ferroelectric domains. *J. Phys. Condens. Matter* **20**, 4 (2008).
8. Nelson, C. T. et al. Spontaneous vortex nanodomain arrays at ferroelectric heterointerfaces. *Nano Lett.* **11**, 828–834 (2011).
9. Dzyaloshinsky, I. Thermodynamic theory of weak ferromagnetism of antiferromagnetics. *J. Phys. Chem. Solids* **4**, 241–255 (1958).
10. Moriya, T. Anisotropic superexchange interaction and weak ferromagnetism. *Phys. Rev.* **120**, 91–98 (1960).
11. Kurumaji, T. et al. Néel-type skyrmion lattice in the tetragonal polar magnet VOSe_2O_5 . *Phys. Rev. Lett.* **119**, 237201–237205 (2017).
12. Kurumaji, T. et al. Direct observation of cycloidal spin modulation and field-induced transition in Néel-type skyrmion-hosting VOSe_2O_5 . *J. Phys. Soc. Jpn* **90**, 024705 (2021).
13. Seddon, S. D. et al. Real-space observation of ferroelectrically induced magnetic spin crystal in SrRuO_3 . *Nat. Commun.* **12**, 2007 (2021).
14. Hong, Z. et al. Stability of polar vortex lattice in ferroelectric superlattices. *Nano Lett.* **17**, 2246–2252 (2017).
15. Damodaran, A. R. et al. Phase coexistence and electric-field control of toroidal order in oxide superlattices. *Nat. Mater.* **16**, 1003–1009 (2017).
16. Stoica, V. A. et al. Optical creation of a supercrystal with three-dimensional nanoscale periodicity. *Nat. Mater.* **18**, 377–383 (2019).

17. Li, X. et al. Atomic-scale observations of electrical and mechanical manipulation of topological polar flux closure. *Proc. Natl Acad. Sci. USA* **117**, 18954–18961 (2020).
18. Ponomareva, I., Naumov, I. I. & Bellaiche, L. Low-dimensional ferroelectrics under different electrical and mechanical boundary conditions: atomistic simulations. *Phys. Rev. B* **72**, 214118 (2005).
19. Sanchez-Santolino, G. et al. Resonant electron tunnelling assisted by charged domain walls in multiferroic tunnel junctions. *Nat. Nanotechnol.* **12**, 655–662 (2017).
20. Penthorn, N. E., Hao, X., Wang, Z., Hua, Y. & Jiang, H. W. Experimental observation of single skyrmion signatures in a magnetic tunnel junction. *Phys. Rev. Lett.* **122**, 257201 (2019).
21. Peters, J. J. P., Apachitei, G., Beanland, R., Alexe, M. & Sanchez, A. M. Polarization curling and flux closures in multiferroic tunnel junctions. *Nat. Commun.* **7**, 13484 (2016).
22. Hadjimichael, M. et al. Metal–ferroelectric supercrystals with periodically curved metallic layers. *Nat. Mater.* **20**, 495–502 (2021).
23. Li, S. et al. Periodic arrays of flux-closure domains in ferroelectric thin films with oxide electrodes. *Appl. Phys. Lett.* **111**, 052901 (2017).
24. Kurumaji, T. et al. Skyrmion lattice with a giant topological Hall effect in a frustrated triangular-lattice magnet. *Science* **365**, 914–918 (2019).
25. Yi, S. D., Onoda, S., Nagaosa, N. & Han, J. H. Skyrmions and anomalous Hall effect in a Dzyaloshinskii-Moriya spiral magnet. *Phys. Rev. B* **80**, 054416–054421 (2009).
26. Zhao, H. J., Chen, P., Prosandeev, S., Artyukhin, S. & Bellaiche, L. Dzyaloshinskii–Moriya-like interaction in ferroelectrics and antiferroelectrics. *Nat. Mater.* **20**, 341–345 (2021).

27. Aguado-Puente, P. & Junquera, J. Ferromagnetic-like closure domains in ferroelectric ultrathin films: first-principles simulations. *Phys. Rev. Lett.* **100**, 177601–177604 (2008).
28. Dürr, H. A. et al. Chiral magnetic domain structures in ultrathin FePd films. *Science* **284**, 2166–2168 (1999).
29. Ahn, C. H. et al. Ferroelectric field effect in ultrathin SrRuO₃ films. *Appl. Phys. Lett.* **70**, 206–208 (1997).
30. Jia, C.-L. et al. Atomic-scale study of electric dipoles near charged and uncharged domain walls in ferroelectric films. *Nat. Mater.* **7**, 57–61 (2008).
31. Naumov, I., Bellaiche, L. & Fu, H. Unusual phase transitions in ferroelectric nanodisks and nanorods. *Nature* **432**, 737–740 (2004).
32. Junquera, J. & Ghosez, P. Critical thickness for ferroelectricity in perovskite ultrathin films. *Nature* **422**, 506–509 (2003).
33. Nahas, Y. et al. Topology and control of self-assembled domain patterns in low-dimensional ferroelectrics. *Nat. Commun.* **11**, 5779 (2020).
34. Nahas, Y. et al. Inverse transition of labyrinthine domain patterns in ferroelectric thin films. *Nature* **577**, 47–51 (2020).
35. Carter, C. B. & Williams, D. B. *Transmission Electron Microscopy: Diffraction, Imaging, and Spectrometry* (Springer, 2016).
36. Bode, M. et al. Chiral magnetic order at surfaces driven by inversion asymmetry. *Nature* **447**, 190–193 (2007).
37. Udalov, O. G., Beloborodov, I. S. & Sapozhnikov, M. V. Magnetic skyrmions and bimerons in films with anisotropic interfacial Dzyaloshinskii-Moriya interaction. *Phys. Rev. B* **103**, 174416 (2021).

38. Chen, J., Zhang, D. W. & Liu, J. M. Exotic skyrmion crystals in chiral magnets with compass anisotropy. *Sci. Rep.* **6**, 29126 (2016).
39. Saw, T. B. et al. Topological defects in epithelia govern cell death and extrusion. *Nature* **544**, 212–216 (2017).
40. Tsesses, S. et al. Optical skyrmion lattice in evanescent electromagnetic fields. *Science* **361**, 993–996 (2018).
41. Giannozzi, P. et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* **29**, 465901 (2017).
42. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
43. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
44. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188 (1976).
45. Zhang, Y., Sun, J., Perdew, J. P. & Wu, X. Comparative first-principles studies of prototypical ferroelectric materials by LDA, GGA, and SCAN meta-GGA. *Phys. Rev. B* **96**, 035143 (2017).
46. Kolpak, A. M., Sai, N. & Rappe, A. Short-circuit boundary conditions in ferroelectric PbTiO₃ thin films. *Phys. Rev. B* **74**, 054112–054117 (2006).
47. Leferovich, R. P. & Mitchell, R. H. A structural study of ternary lanthanide orthoscandate perovskites. *J. Solid State Chem.* **177**, 2188–2197 (2004).
48. Longo, V. M. et al. On the photoluminescence behavior of samarium-doped strontium titanate nanostructures under UV light. A structural and electronic understanding. *Phys. Chem. Chem. Phys.* **12**, 7566–7579 (2010).

49. Bansal, C., Kawanaka, H., Takahashi, R. & Nishihara, Y. Metal–insulator transition in Fe-substituted SrRuO₃ bad metal system. *J. Alloys Compd.* **360**, 47–53 (2003).
50. Kuroiwa, Y., Aoyagi, S. & Sawada, A. Evidence for Pb-O covalency in tetragonal PbTiO₃. *Phys. Rev. Lett.* **87**, 217601–217605 (2001).
51. Jones, L. et al. Smart Align—a new tool for robust non-rigid registration of scanning microscope data. *Adv. Struct. Chem. Imaging* **1**, 8 (2015).
52. Li, Q. et al. Quantification of flexoelectricity in PbTiO₃/SrTiO₃ superlattice polar vortices using machine learning and phase-field modeling. *Nat. Commun.* **8**, 1468 (2017).
53. Aso, R., Kan, D., Shimakawa, Y. & Kurata, H. Octahedral tilt propagation controlled by A-site cation size at perovskite oxide heterointerfaces. *Cryst. Growth Des.* **14**, 2128–2132 (2014).
54. Peters, J. J. P. et al. Polarization screening mechanisms at La_{0.7}Sr_{0.3}MnO₃–PbTiO₃ interfaces. *ACS Appl. Mater. Interfaces* **12**, 10657–10663 (2020).
55. Sanchez-Santolino, G. et al. Oxygen octahedral distortions in LaMO₃/SrTiO₃ superlattices. *Microsc. Microanal.* **20**, 825–831 (2014).
56. Rondinelli, J. M. & Spaldin, N. A. Structure and properties of functional oxide thin films: insights from electronic-structure calculations. *Adv. Mater.* **23**, 3363–3381 (2011).
57. Sepliarsky, M., Stachiotti, M. G. & Migoni, R. L. Surface reconstruction and ferroelectricity in PbTiO₃ thin films. *Phys. Rev. B* **72**, 014110–014116 (2005).
58. Zhang, S. et al. Polarization rotation in ultrathin ferroelectrics tailored by interfacial oxygen octahedral coupling. *ACS Nano* **12**, 3681–3688 (2018).

Acknowledgements

This work was partly supported by the EPSRC (UK) through grant nos. EP/P031544/1 and EP/P025803/1. M.A. acknowledges the Theo Murphy Blue Skies Award of the Royal Society. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science user facility, operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. We also acknowledge the technical support from M. Crosbie. We would like to acknowledge the University of Warwick Research Technology Platform for assistance in the research described in this paper.

Author information

Affiliations

1. Department of Physics, University of Warwick, Coventry, UK

Dorin Rusu, Jonathan J. P. Peters, Thomas P. A. Hase, James A. Gott, Samuel D. Seddon, Richard Beanland, Ana M. Sanchez & Marin Alexe

2. School of Physics, Trinity College Dublin, Dublin, Ireland

Jonathan J. P. Peters

3. Diamond Light Source, Didcot, UK

Gareth A. A. Nisbet

4. Argonne National Laboratory, Lemont, IL, USA

Jörg Strempfer & Daniel Haskel

Contributions

M.A. conceived the idea. D.R., M.A. and A.M.S. designed the experiments. D.R. prepared the samples, performed DFT and DF-TEM experiments and analysed the data. J.J.P.P. and J.A.G. performed the STEM experiments and analysis. G.A.A.N., J.S., D.H. and D.R. collected the synchrotron data. T.P.A.H. and D.R. analysed the XRD data. R.B. performed the two-beam diffraction contrast simulations. All authors contributed to the discussions. All authors wrote the manuscript.

Corresponding author

Correspondence to [Marin Alexe](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

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 Structure characterization.](#)

a, AFM topography of the surface of our $(\text{SrRuO}_3)_{11}/(\text{PbTiO}_3)_{13}/(\text{SrRuO}_3)_{11}$ trilayer sample. **b**, Low-magnification cross-sectional STEM of the sample. The scale bar is 10 nm.

Extended Data Fig. 2 RSM data.

a, 3D reciprocal space map around the DSO 002_{pc} Bragg peak. The side panels show the 2D projected intensity of the recorded scatter. **b**, RSM around the asymmetric reflection $(103)_{\text{pc}}$. **c**, Reciprocal space map Q_z versus Q_x . **d**, Comparison plot of the extracted Q_z scan at $Q_x = 0$ and $Q_x = 0.07 \text{ \AA}^{-1}$. **e**, Integrated (boxed area) line profile showing the first-order and second-order satellite peaks and their widths. **f**, Reciprocal space map Q_z versus Q_y . **g**, The integrated (boxed area) line profile shows weak first-order satellite peaks corresponding to a periodicity of about 8.05 nm. **h**, Plan-view projection into a Q_x versus Q_y RSM map with extracted line scans showing the in-plane distribution of the satellite peaks.

Extended Data Fig. 3 Cross-sectional DF-TEM.

a, Image of a $(100)_{\text{pc}}$ cross section taken under the $\mathbf{g} = 020_{\text{pc}}$ excitation condition. **b**, Image of the same $(100)_{\text{pc}}$ cross section taken under the $\mathbf{g} = 002_{\text{pc}}$ excitation. **c**, Image of a $(010)_{\text{pc}}$ cross section (that is, cut at 90° from **a** and **b**) taken using $\mathbf{g} = 002_{\text{pc}}$. The scale bars are 20 nm.

Extended Data Fig. 4 Plan-view DF-TEM.

a, Low-magnification plan view of the complex domain pattern take under the $\mathbf{g}_1 = 110_{\text{pc}}$ condition. The figure inset shows the enlarged boxed area. **b**, Plan-view dark-field image taken under $\mathbf{g}_2 = 100_{\text{pc}}$. **c**, Plan-view diffraction contrast taken under $\mathbf{g}_3 = 010_{\text{pc}}$ excitation. The scale bars are 100 nm for **a** and 30 nm for **b** and **c**.

Extended Data Fig. 5 Noise filter.

a, 2D Fourier transform of the plan-view image taken under $\mathbf{g}_1 = 110$ excitation. **b**, Bandpass filter that removes the noise and retains the signal for $|\mathbf{Q}| < 0.16 \text{ nm}^{-1}$. **c**, Bandpass filter that also removes the central spot and the signal for $|\mathbf{Q}| < 0.015 \text{ nm}^{-1}$. The scale bar is 30 nm.

Extended Data Fig. 6 Filtered/unfiltered plan-view DF-TEM.

a, The bandpass filter improves the signal-to-noise ratio without introducing artefacts. Both the filtered and unfiltered images show, apart from the labyrinth pattern, a periodic modulation in the contrast along the individual domains. **b**, The second modulation permeates the labyrinth pattern.

Extended Data Fig. 7 Diffraction contrast simulations.

Left, experimental plan-view DF-TEM images of the vortex array. Note that the $\mathbf{g} = 110_{\text{pc}}$ image is at higher magnification and Bragg filtered. Right, two-beam Howie–Whelan diffraction contrast simulations of the contrast arising from the strain fields of a 2D array of vortices as described in the main text (deviation parameter $s = 0.01 \text{ nm}$). The scale bars are 30 nm for $\mathbf{g} = 010$, 30 nm for $\mathbf{g} = 100$ and 10 nm for $\mathbf{g} = 110$.

Extended Data Fig. 8 Tilt map.

Left, oxygen tilt behaviour along a row of unit cells. Right, tilt map throughout the PTO layer.

Extended Data Fig. 9 Cross-sectional polar maps.

a, Polarization maps along the ordered $[010]_{\text{pc}}$ direction. The projection of a cycloidal and helical modulated vortex array into the $[001]_{\text{pc}}-[010]_{\text{pc}}$ plane shows that the domain topology is retained. **b**, The projection of the modulated vortex array into the $[001]_{\text{pc}}-[100]_{\text{pc}}$ plane shows that the cycloidal modulation allows the polar vector to rotate in plane, similar to the experimental polar map, whereas the helical modulation does not.

Extended Data Fig. 10 X-ray CD data.

a, $[100]_{\text{pc}}//[001]_o$ in the scattering plane. **b**, $[010]_{\text{pc}}//[-110]_o$ in the scattering plane. The first row presents the sum of the dichroic signal, $(I^+ + I^-)/2$, in which I^\pm refers to the measured intensity for opposite helicities of the incoming light and the second row shows the CD signal, $(I^+ - I^-)/(I^+ + I^-)$, and its behaviour on 180° rotation of the sample. The third panel shows the dichroic signal, $(CD_{\phi 1} - CD_{\phi 2})/2$, associated with a rotation of the sample by 180° , demonstrating a weak signal at the \pm satellites in **a**, which is absent in **b**. In **c** and **d**, we show the dichroism under sample rotation when $[100]_{\text{pc}}//[001]_o$ is in the scattering plane (**c**) and when the $[010]_{\text{pc}}//[-110]_o$ direction is in the scattering plane (**d**). The upper panels of **c** and **d** plot the data from the two satellites onto a common axis, with the lower panels showing their average.

Supplementary information

Supplementary Information

This file contains Supplementary Text, Supplementary Figures 1–10 and Supplementary References.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Rusu, D., Peters, J.J.P., Hase, T.P.A. *et al.* Ferroelectric incommensurate spin crystals. *Nature* **602**, 240–244 (2022). <https://doi.org/10.1038/s41586-021-04260-1>

- Received: 08 March 2021
- Accepted: 17 November 2021
- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04260-1>

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/s41586-021-04260-1>

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

- Article
- [Published: 09 February 2022](#)

Time-reversal symmetry-breaking charge order in a kagome superconductor

- [C. Mielke III](#) ORCID: orcid.org/0000-0001-7773-9485^{1,2 na1},
- [D. Das](#) ORCID: orcid.org/0000-0001-5453-0195^{1 na1},
- [J.-X. Yin](#)^{3 na1},
- [H. Liu](#)^{4,5 na1},
- [R. Gupta](#)¹,
- [Y.-X. Jiang](#)³,
- [M. Medarde](#)⁶,
- [X. Wu](#)⁷,
- [H. C. Lei](#) ORCID: orcid.org/0000-0003-0850-8514⁸,
- [J. Chang](#)²,
- [Pengcheng Dai](#)⁹,
- [Q. Si](#) ORCID: orcid.org/0000-0003-1357-2705⁹,
- [H. Miao](#)¹⁰,
- [R. Thomale](#) ORCID: orcid.org/0000-0002-3979-8836^{11,12},
- [T. Neupert](#) ORCID: orcid.org/0000-0003-0604-041X²,
- [Y. Shi](#)^{4,5,13},
- [R. Khasanov](#) ORCID: orcid.org/0000-0002-4768-5524¹,
- [M. Z. Hasan](#) ORCID: orcid.org/0000-0001-9730-3128^{3,14,15,16},
- [H. Luetkens](#) ORCID: orcid.org/0000-0002-5842-9788¹ &
- [Z. Guguchia](#) ORCID: orcid.org/0000-0002-5498-328X¹

[Nature](#) volume **602**, pages 245–250 (2022)

- 1960 Accesses
- 91 Altmetric
- [Metrics details](#)

Subjects

- [Magnetic properties and materials](#)
- [Superconducting properties and materials](#)

Abstract

The kagome lattice¹, which is the most prominent structural motif in quantum physics, benefits from inherent non-trivial geometry so that it can host diverse quantum phases, ranging from spin-liquid phases, to topological matter, to intertwined orders^{2,3,4,5,6,7,8} and, most rarely, to unconventional superconductivity^{6,9}. Recently, charge sensitive probes have indicated that the kagome superconductors AV_3Sb_5 ($A = K, Rb, Cs$)^{9,10,11} exhibit unconventional chiral charge order^{12,13,14,15,16,17,18,19}, which is analogous to the long-sought-after quantum order in the Haldane model²⁰ or Varma model²¹. However, direct evidence for the time-reversal symmetry breaking of the charge order remains elusive. Here we use muon spin relaxation to probe the kagome charge order and superconductivity in KV_3Sb_5 . We observe a noticeable enhancement of the internal field width sensed by the muon ensemble, which takes place just below the charge ordering temperature and persists into the superconducting state. Notably, the muon spin relaxation rate below the charge ordering temperature is substantially enhanced by applying an external magnetic field. We further show the multigap nature of superconductivity in KV_3Sb_5 and that the $\langle T \rangle / \lambda_{ab}^{2}$ ratio (where T_c is the superconducting transition temperature and λ_{ab} is the magnetic penetration depth in the kagome plane) is comparable to those of unconventional high-temperature superconductors. Our results point to time-reversal symmetry-

breaking charge order intertwining with unconventional superconductivity in the correlated kagome lattice.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Indication of time-reversal symmetry-breaking of the charge order in KV₃Sb₅.

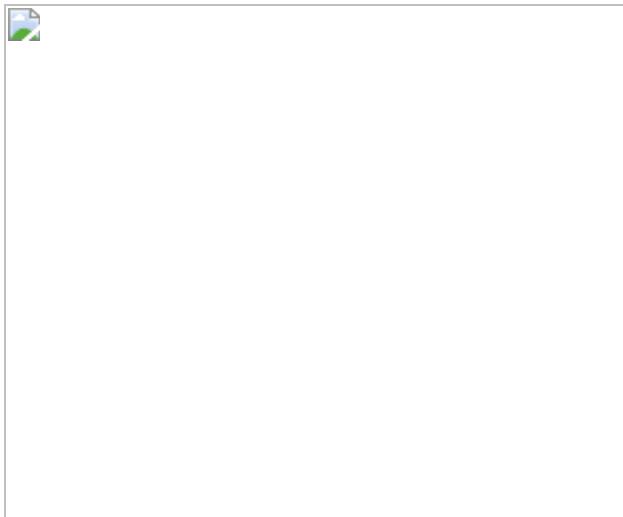


Fig. 2: Enhanced magnetic response of the charge order with applying external magnetic fields.

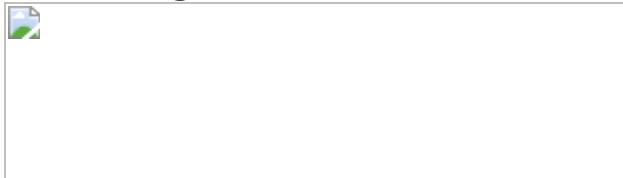
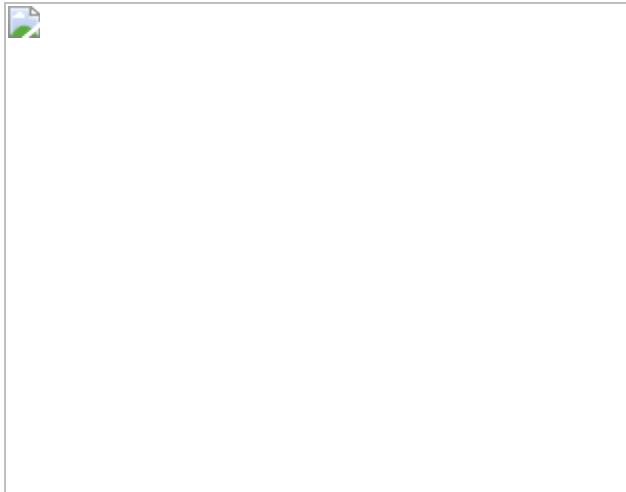


Fig. 3: Correlated kagome superconductivity.



Data availability

All relevant data are available from the authors. Alternatively, the data can be accessed through the data base at the following link:
<http://musruser.psi.ch/cgi-bin/SearchDB.cgi>.

References

1. Syôzi, I. Statistics of kagome lattice. *Prog. Theor. Phys.* **6**, 306-308 (1951).
2. Barz, H. Ternary transition metal phosphides: high-temperature superconductors. *Mater. Res. Bull.* **15**, 1489-1491 (1980).
3. Zhou, Y., Kanoda, K. & Ng, T.-K. Quantum spin liquid states. *Rev. Mod. Phys.* **89**, 025003 (2017).
4. Guguchia, Z. et al. Tunable anomalous Hall conductivity through volume-wise magnetic competition in a topological kagome magnet. *Nat. Commun.* **11**, 559 (2020).
5. Yin, J.-X., Pan, S. H. & Hasan, M. Z. Probing topological quantum matter with scanning tunneling microscopy. *Nat. Rev. Phys.* **3**, 249–263 (2021).

6. Mielke, III, C. et al. Nodeless kagome superconductivity in LaRu_3Si_2 . *Phys. Rev. Materials* **5**, 034803 (2021).
7. J.-X. Yin, et. al., Quantum-limit Chern topological magnetism in TbMn_6Sn_6 . *Nature* **583**, 533–536 (2020).
8. Pershoguba, S. S. & Yakovenko, V. M. Optical control of topological memory based on orbital magnetization. Preprint at <http://arxiv.org/abs/2106.01192> (2021).
9. Ortiz, B. et al. CsV_3Sb_5 : a Z_2 topological kagome metal with a superconducting ground state. *Phys. Rev. Lett.* **125**, 247002 (2020).
10. Ortiz, B. et al. Superconductivity in the Z_2 kagome metal KV_3Sb_5 . *Phys. Rev. Materials* **5**, 034801 (2021).
11. Yin, Q. et al. Superconductivity and normal-state properties of kagome metal RbV_3Sb_5 single crystals. *Chin. Phys. Lett.* **38**, 037403 (2021).
12. Jiang, Y.-X. et al. Discovery of topological charge order in kagome superconductor KV_3Sb_5 . *Nat. Mater.* **20**, 1353–1357 (2021).
13. Shumiya, N. et al. Tunable chiral charge order in kagome superconductor RbV_3Sb_5 . *Phys. Rev. B* **104**, 035131 (2021).
14. Wang, Z. et al. Electronic nature of chiral charge order in the kagome superconductor CsV_3Sb_5 . *Phys. Rev. B* **104**, 075148 (2021).
15. Feng, X., Jiang, K., Wang, Z. & Hu, J. Chiral flux phase in the Kagome superconductor $AV_3\text{Sb}_5$. *Sci. Bull.* **66**, 1384–1388 (2021).
16. Denner, M., Thomale, R. & Neupert, T. Analysis of charge order in the kagome metal $AV_3\text{Sb}_5$ ($A = \text{K}, \text{Rb}, \text{Cs}$). *Phys. Rev. Lett.* **127**, 217601 (2021).
17. Lin, Y.-P. & Nandkishore, R. Complex charge density waves at Van Hove singularity on hexagonal lattices: Haldane-model phase diagram

- and potential realization in kagome metals AV_3Sb_5 . *Phys. Rev. B* **104**, 045122 (2021).
18. Wu, X. et al. Nature of unconventional pairing in the kagome superconductors AV_3Sb_5 . *Phys. Rev. Lett.* **127**, 177001 (2021).
 19. Setty, C., Hu, H., Chen, L. & Si, Q. Electron correlations and T -breaking density wave order in a Z_2 kagome metal. Preprint at <https://arxiv.org/abs/2105.15204> (2021).
 20. Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: condensed-matter realization of the parity anomaly. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
 21. Varma, C. M. Non-Fermi-liquid states and pairing instability of a general model of copper oxide metals. *Phys. Rev. B* **55**, 14554–14580 (1997).
 22. Chakravarty, S., Laughlin, R., Morr, D. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
 23. Yang, S. et al. Giant, unconventional anomalous Hall effect in the metallic frustrated magnet candidate, KV_3Sb_5 . *Sci. Adv.* **6**, eabb6003 (2020).
 24. Sonier, J. E., Brewer, J. H. & Kiefl, R. F. μ SR studies of the vortex state in type-II superconductors. *Rev. Mod. Phys.* **72**, 769 (2000).
 25. Luke, G. M. et al. Time-reversal symmetry breaking superconductivity in Sr_2RuO_4 . *Nature* **394**, 558-561 (1998).
 26. Kenney, E., Ortiz, B., Wang, C., Wilson, S. & Graf, M. Absence of local moments in the kagome metal KV_3Sb_5 as determined by muon spin spectroscopy. *J. Phys. Condens. Matter* **33**, 235801 (2021).
 27. Kubo, R. & Toyabe, T. *Magnetic Resonance and Relaxation* (North Holland, 1967).

28. Huang, W. et al. Precision search for magnetic order in the pseudogap regime of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ by muon spin relaxation. *Phys. Rev. B* **85**, 104527 (2012).
29. Singh, A. D. et al. Time-reversal symmetry breaking and multigap superconductivity in the noncentrosymmetric superconductor La_7Ni_3 . *Phys. Rev. B* **103**, 174502 (2021).
30. Sedlak, K., Scheuermann, R., Stoykov, A. & Amato, A. GEANT4 simulation and optimisation of the high-field μSR spectrometer. *Physica B* **404**, 970–973 (2009).
31. Xiang, Y. et al. Twofold symmetry of c -axis resistivity in topological kagome superconductor CsV_3Sb_5 with in-plane rotating magnetic field. Preprint at <https://arxiv.org/abs/2104.06909> (2021).
32. Zhao, H. et al. Cascade of correlated electron states in a kagome superconductor CsV_3Sb_5 . Preprint at <https://arxiv.org/pdf/2103.03118.pdf> (2021).
33. Khasanov, R. et al. Evolution of two-gap behavior of the superconductor FeSe_{1-x} . *Phys. Rev. Lett.* **104**, 087004 (2010).
34. Kogan, V. G., Martin, C. & Prozorov, R. Superfluid density and specific heat within a self-consistent scheme for a two-band superconductor. *Phys. Rev. B* **80**, 014507 (2009).
35. Gupta, R. et al. Microscopic evidence for anisotropic multigap superconductivity in the CsV_3Sb_5 kagome superconductor. Preprint at <https://arxiv.org/abs/2108.01574> (2021).
36. Han-Shu, X. et al. Multiband superconductivity with sign-preserving order parameter in kagome superconductor CsV_3Sb_5 . Preprint at <https://arxiv.org/pdf/2104.08810.pdf> (2021).
37. Uemura, Y. J. et al. Universal correlations between T_c and n_s/m^* (carrier density over effective mass) in high- T_c cuprate

- superconductors. *Phys. Rev. Lett.* **62**, 2317 (1989).
38. von Rohr, F. O. et al. Unconventional scaling of the superfluid density with the critical temperature in transition metal dichalcogenides. *Sci. Adv.* **5**, eaav8465 (2019).
39. Shengelaya, A. et al. Muon-spin-rotation measurements of the penetration depth of the infinite-layer electron-doped $\text{Sr}_{0.9}\text{La}_{0.1}\text{CuO}_2$ cuprate superconductor. *Phys. Rev. Lett.* **94**, 127001 (2005).
40. Luetkens, H. et al. The electronic phase diagram of the $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ superconductor. *Nat. Mater.* **8**, 305-309 (2009).
41. Amato, A. et al. The new versatile general purpose surface-muon instrument (GPS) based on silicon photomultipliers for μSR measurements on a continuous-wave beam. *Rev. Sci. Instrum.* **88**, 093301 (2017).
42. Suter, A. and Wojek, B. M. Musrfit: a free platform-independent framework for μSR data analysis. *Phys. Procedia* **30**, 69 (2012).
43. Kiesel, M. L. & Thomale, R. Sublattice interference in the kagome Hubbard model. *Phys. Rev. B* **86**, 121105 (2012).
44. Lee, S. L. et al. Evidence for two-dimensional thermal fluctuations of the vortex structure in $\text{Bi}_{2.15}\text{Sr}_{1.85}\text{CaCu}_2\text{O}_{8+\Delta}$ from muon spin rotation experiments. *Phys. Rev. Lett.* **75**, 922 (1995).
45. Guguchia, Z. et al. Signatures of the topological s^{+-} superconducting order parameter in the type-II Weyl semimetal $\text{T}_d\text{-MoTe}_2$. *Nat. Commun.* **8**, 1082 (2017).
46. Brandt, E. H. Flux distribution and penetration depth measured by muon spin rotation in high- T_c superconductors. *Phys. Rev. B* **37**, 2349 (1988).

47. Bouquet, F. et al. Phenomenological two-gap model for the specific heat of MgB₂. *Europhys. Lett.* **56**, 856 (2001).
48. Prozorov, R. & Giannetta, R. W. Magnetic penetration depth in unconventional superconductors. *Supercond. Sci. Technol.* **19**, R41 (2006).
49. Khasanov, R. et al. Experimental evidence for two gaps in the high-temperature La_{1.83}Sr_{0.17}CuO₄ superconductor. *Phys. Rev. Lett.* **98**, 057007 (2007).
50. Khasanov, R. et al. SrPt₃P: a two-band single-gap superconductor. *Phys. Rev. B* **90**, 140507(R) (2014).
51. Kogan, V. G. London approach to anisotropic type-II superconductors. *Phys. Rev. B* **24**, 1572 (1981).
52. Gupta, R. et al. Self-consistent two-gap approach in studying multi-band superconductivity in NdFeAsO_{0.65}F_{0.35}. *Front. Phys.* **8**, 2 (2020).

Acknowledgements

The μSR experiments were carried out at the Swiss Muon Source (SμS) Paul Scherrer Institute, Villigen, Switzerland. The magnetization measurements were carried out on the MPMS device of the Laboratory for Multiscale Materials Experiments, Paul Scherrer Institute, Villigen, Switzerland (SNSF grant no. 206021_139082). Z.G. acknowledges the useful discussions with R. Scheuermann and A. Amato. Z.G., C.M.III and D.D. thank C. Baines for the technical assistance during DOLLY experiments. M.Z.H. acknowledges visiting scientist support from IQIM at the California Institute of Technology. Experimental and theoretical work at Princeton University was supported by the Gordon and Betty Moore Foundation (GBMF4547 and GBMF9461; M.Z.H.) and the material characterization is supported by the US Department of Energy under the Basic Energy Sciences programme (grant no. DOE/BES DE-FG-02-05ER46200). The theory work at Rice has primarily been supported by the US Department of Energy, BES under award no. DE-SC0018197 and has

also been supported by the Robert A. Welch Foundation grant no. C-1411 (Q.S.). The work at Rice university is also supported by US Department of Energy, BES under Grant No. DE-SC0012311 (P.D.). This work is also supported by the Beijing Natural Science Foundation (grant no. Z180008, Z200005), the National Key Research and Development Program of China (grant no. 2017YFA0302900, Y2018YFE0202600) and the National Natural Science Foundation of China (grant no. U2032204). The work of R.G. was supported by the Swiss National Science Foundation (SNF grant no. 200021_175935). T.N. acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (ERC-StG-Neupert-757867-PARATOP). H. M. was sponsored by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. R.T. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Project-ID 258499086-SFB1170 and by the Wurzburg-Dresden Cluster of Excellence on Complexity and Topology in Quantum Matter-ct.qmat Project-ID 390858490-EXC 2147.

Author information

Author notes

1. These authors contributed equally: C. Mielke III, D. Das, J.-X. Yin, H. Liu

Affiliations

1. Laboratory for Muon Spin Spectroscopy, Paul Scherrer Institute, Villigen PSI, Switzerland

C. Mielke III, D. Das, R. Gupta, R. Khasanov, H. Luetkens & Z. Guguchia

2. Physik-Institut, Universität Zürich, Winterthurerstrasse 190, Zürich, Switzerland

C. Mielke III, J. Chang & T. Neupert

3. Laboratory for Topological Quantum Matter and Advanced Spectroscopy (B7), Department of Physics, Princeton University, Princeton, NJ, USA

J.-X. Yin, Y.-X. Jiang & M. Z. Hasan

4. Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China

H. Liu & Y. Shi

5. School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China

H. Liu & Y. Shi

6. Laboratory for Multiscale Materials Experiments, Paul Scherrer Institut, Villigen PSI, Switzerland

M. Medarde

7. Max-Planck-Institut für Festkörperforschung, Stuttgart, Germany

X. Wu

8. Department of Physics and Beijing Key Laboratory of Opto-electronic Functional Materials and Micro-nano Devices, Renmin University of China, Beijing, China

H. C. Lei

9. Department of Physics and Astronomy, Rice Center for Quantum Materials, Rice University, Houston, TX, USA

Pengcheng Dai & Q. Si

10. Material Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

H. Miao

11. Institut fur Theoretische Physik und Astrophysik, Universitat Wurzburg, Wurzburg, Germany

R. Thomale

12. Department of Physics, Indian Institute of Technology Madras, Chennai, India

R. Thomale

13. Songshan Lake Materials Laboratory, Dongguan, Guangdong, China

Y. Shi

14. Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ, USA

M. Z. Hasan

15. Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

M. Z. Hasan

16. Quantum Science Center, Oak Ridge, TN, USA

M. Z. Hasan

Contributions

Z.G., Y.-X.J. and M.Z.H. conceived the study. Z.G. supervised the project. Sample growth and single-crystal X-ray diffraction experiments were carried out by H. Liu and Y.S. Magnetization and Laue X-ray diffraction experiments were performed by C.M.III, Z.G., M.M. and H.C.L. μ SR

experiments and corresponding discussions were carried out by Z.G., C.M.III, D.D., R.G., R.K., H.Luet., J.J.C., J.-X.Y., Y.-X.J., M.Z.H., X.W., P.D., Q.S., H.M., R.T. and T.N. μ SR data analysis was undertaken by Z.G. and C.M.III, with contributions from R.K., H.L., D.D. and R.G. STM experiments were performed by J.-X.Y., Y.-X.J. and M.Z.H. Figure development and the writing of the paper were carried out by Z.G. and C.M.III, with contributions from J.-X.Y., H.Luet. and M.Z.H. All authors discussed the results, interpretation and conclusion.

Corresponding authors

Correspondence to [M. Z. Hasan](#) or [Z. Guguchia](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Victor Yakovenko and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 Crystal structure of KV₃Sb₅.](#)

Three dimensional representation (**a**) and top view (**b**) of the atomic structure of KV_3Sb_5 . In panel (**c**) is displayed an optical microscope image of a $3 \times 2 \times 0.2$ mm single crystal of KV_3Sb_5 on millimeter paper, with the scale shown. The hexagonal symmetry is immediately apparent. (**d**) Scanning Transmission Microscope (STM) image of the V kagome lattice from a cryogenically cleaved sample.

Extended Data Fig. 2 Single Crystal X-Ray Diffraction for KV_3Sb_5 .

(**a**) X-ray diffraction image for KV_3Sb_5 recorded at 300 K. The well-defined peaks are labeled with their crystallographic indices. No second phase has been detected. (**b**) Laue X-ray diffraction image of the single crystal sample KV_3Sb_5 , oriented with the *c*-axis along the beam. (**c**) The temperature dependence of magnetic susceptibility of KV_3Sb_5 above 1.8 K. It shows an anomaly at $T^* \approx 80$ K, coinciding with emergence of a charge order.

Extended Data Fig. 3 Anisotropic magnetic response across charge order temperature in the single crystalline sample of KV_3Sb_5 .

(**a**) The temperature dependence of magnetic susceptibility for KV_3Sb_5 measured at various magnetic fields applied parallel to the *c*-axis. (**b**) The temperature dependence of magnetic susceptibility for KV_3Sb_5 measured in the field of 1 T, applied both parallel to the kagome plane and parallel to the *c*-axis.

Extended Data Fig. 4 Zero-field μSR experiment for the single crystalline sample of KV_3Sb_5 .

The ZF μSR time spectra for KV_3Sb_5 , obtained at $T = 5$ K from detectors 3 & 4 and 2 & 1. The solid curves represent fits to the recorded time spectra, using only Gaussian Kubo Toyabe (GKT) function (red) and the one with

an additional exponential $\exp(-\Gamma t)$ term (blue). The inset shows the low time part of the spectrum.

Extended Data Fig. 5 Zero-field μ SR experiment for the polycrystalline sample of KV₃Sb₅.

The ZF μ SR time spectra for the polycrystalline sample of KV₃Sb₅, obtained at $T = 5$ K. The solid curves represent fits to the recorded time spectra, using only Gaussian Kubo Toyabe (GKT) function (red) and the one with an additional exponential $\exp(-\Gamma t)$ term (blue). The inset shows the low time part of the spectrum.

Extended Data Fig. 6 High-field μ SR experiment for KV₃Sb₅.

Fourier transform for the μ SR asymmetry spectra of KV₃Sb₅ at 5 K for the applied field of $\mu_0 H = 8$ T. The black solid line represents the fit to the data using the two component signal. Red and blue solid lines show the signals arising from the sample and the silver sample holder (majority), respectively. The inset shows the temperature dependences of the muon spin relaxation rates arising from the sample and the silver sample holder.

Extended Data Fig. 7 Superconducting gap symmetry in KV₃Sb₅.

- (a) The SC muon depolarization rates $\sigma_{SC,ab}$, and $\sigma_{SC,ac}$ as well as the inverse squared magnetic penetration depth λ_{ab}^{-2} and λ_{ac}^{-2} as a function of temperature, measured in 5 mT, applied parallel and perpendicular to the kagome plane. (b) The SC muon depolarization rate $\sigma_{SC,ac}$, measured in 10 mT, applied parallel to the kagome plane. The solid line represents the indistinguishable 2-gap *s*-wave and *s+d* wave model. The error bars represent the s.d. of the fit parameters. (c) Temperature dependence of the difference between the internal field $\mu_0 H_{SC}$ measured in the SC state and the one measured in the normal state $\mu_0 H_{NS}$ at $T = 5$ K for KV₃Sb₅.

Extended Data Fig. 8 A self-consistent approach for a two-band superconductor in KV₃Sb₅.

The SC muon depolarization rates $\sigma_{SC,c}$ (**a**), and $\sigma_{SC,ab}$ (**b**) as a function of temperature, measured in 5 mT, applied perpendicular and parallel to the kagome plane. (**c**) The SC muon depolarization rate $\sigma_{SC,ac}$, measured in 10 mT, applied parallel to the kagome plane. The solid black and purple lines are the theoretical curves obtained within the framework of self-consistent approach for a two-band superconductor described in the text. The red and the blue dashed lines correspond to the contribution of the large and the small superconducting gaps to the total superfluid density, solid black lines. The insets show the temperature dependences of the large Δ_1 and the small Δ_2 .

Extended Data Table 1 Atomic positions

Extended Data Table 2 Crystallographic refinement

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Mielke, C., Das, D., Yin, JX. *et al.* Time-reversal symmetry-breaking charge order in a kagome superconductor. *Nature* **602**, 245–250 (2022). <https://doi.org/10.1038/s41586-021-04327-z>

- Received: 25 June 2021
- Accepted: 07 December 2021
- Published: 09 February 2022

- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04327-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

[Electrons go loopy in a family of superconductors](#)

- Morten H. Christensen
- Turan Birol

News & Views 09 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04327-z>

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

- Article
- [Published: 09 February 2022](#)

A highly distorted ultraelastic chemically complex Elinvar alloy

- [Q. F. He](#) [ORCID: orcid.org/0000-0002-1728-8579](#)¹ na1,
- [J. G. Wang](#)^{1,2} na1,
- [H. A. Chen](#) [ORCID: orcid.org/0000-0002-0728-2165](#)³ na1,
- [Z. Y. Ding](#)¹,
- [Z. Q. Zhou](#)¹,
- [L. H. Xiong](#)⁴,
- [J. H. Luan](#)⁵,
- [J. M. Pelletier](#)⁶,
- [J. C. Qiao](#)^{1,6,7},
- [Q. Wang](#)⁸,
- [L. L. Fan](#)⁹,
- [Y. Ren](#) [ORCID: orcid.org/0000-0001-9831-6035](#)^{4,10},
- [Q. S. Zeng](#) [ORCID: orcid.org/0000-0001-5960-1378](#)^{11,12},
- [C. T. Liu](#) [ORCID: orcid.org/0000-0001-7888-9725](#)^{1,5},
- [C. W. Pao](#) [ORCID: orcid.org/0000-0003-0821-7856](#)³,
- [D. J. Srolovitz](#) [ORCID: orcid.org/0000-0001-6038-020X](#)^{13,14} &
- [Y. Yang](#) [ORCID: orcid.org/0000-0002-0491-8295](#)^{1,5}

[Nature](#) volume **602**, pages 251–257 (2022)

- 1879 Accesses
- 92 Altmetric
- [Metrics details](#)

Subjects

- [Mechanical properties](#)
- [Metals and alloys](#)

Abstract

The development of high-performance ultraelastic metals with superb strength, a large elastic strain limit and temperature-insensitive elastic modulus (Elinvar effect) are important for various industrial applications, from actuators and medical devices to high-precision instruments^{1,2}. The elastic strain limit of bulk crystalline metals is usually less than 1 per cent, owing to dislocation easy gliding. Shape memory alloys³—including gum metals^{4,5} and strain glass alloys^{6,7}—may attain an elastic strain limit up to several per cent, although this is the result of pseudo-elasticity and is accompanied by large energy dissipation³. Recently, chemically complex alloys, such as ‘high-entropy’ alloys⁸, have attracted tremendous research interest owing to their promising properties^{9,10,11,12,13,14,15}. In this work we report on a chemically complex alloy with a large atomic size misfit usually unaffordable in conventional alloys. The alloy exhibits a high elastic strain limit (approximately 2 per cent) and a very low internal friction (less than 2×10^{-4}) at room temperature. More interestingly, this alloy exhibits an extraordinary Elinvar effect, maintaining near-constant elastic modulus between room temperature and 627 degrees Celsius (900 kelvin), which is, to our knowledge, unmatched by the existing alloys hitherto reported.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Structure characterization of single-crystal $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.

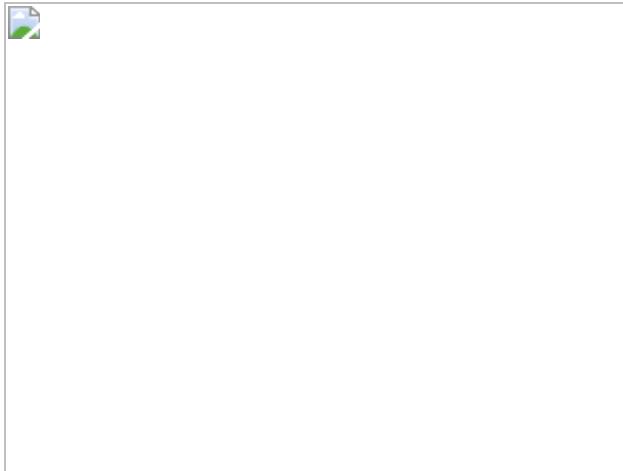


Fig. 2: DFT calculation of three structure models for the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.

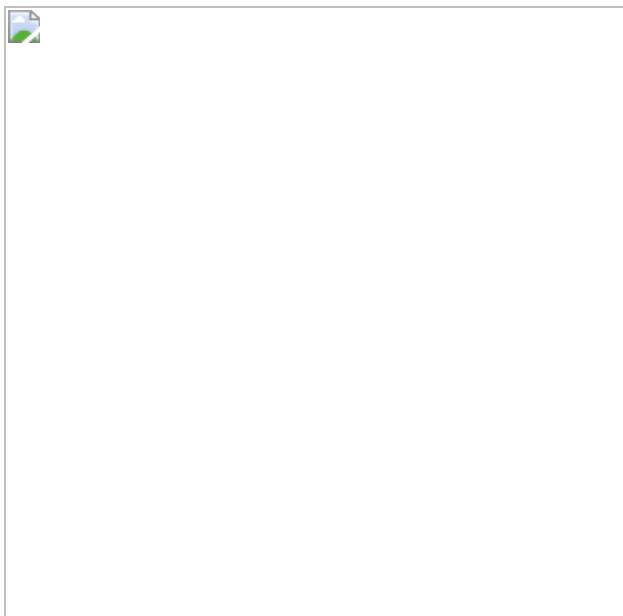
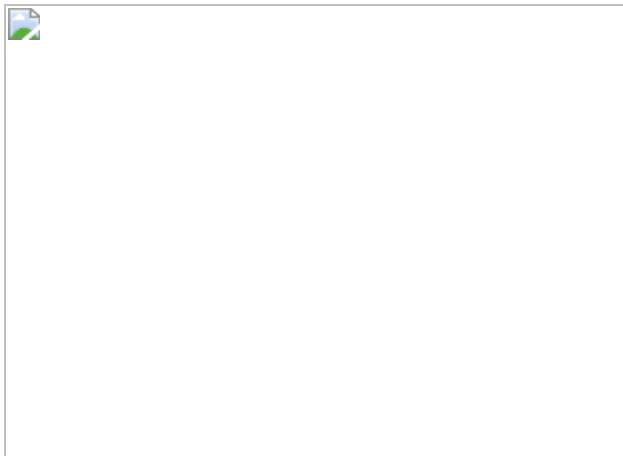


Fig. 3: Mechanical properties of the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.



Fig. 4: The Elinvar effect in the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.



Data availability

The data supporting the findings of this study are available within the article.

References

1. Szold, A. Nitinol: shape-memory and super-elastic materials in surgery. *Surg. Endosc.* **20**, 1493–1496 (2006).
2. Duerig, T. W. Present and future applications of shape memory and superelastic materials. *MRS Proc.* **360**, 497 (2011).

3. Tanaka, Y. et al. Ferrous polycrystalline shape-memory alloy showing huge superelasticity. *Science* **327**, 1488–1490 (2010).
4. Saito, T. et al. Multifunctional alloys obtained via a dislocation-free plastic deformation mechanism. *Science* **300**, 464–467 (2003).
5. Gutkin, M. Yu., Ishizaki, T., Kuramoto, S. & Ovid'ko, I. A. Nanodisturbances in deformed gum metal. *Acta Mater.* **54**, 2489–2499 (2006).
6. Wang, Y., Ren, X. & Otsuka, K. Shape memory effect and superelasticity in a strain glass alloy. *Phys. Rev. Lett.* **97**, 225703 (2006).
7. Ren, X. Strain glass and ferroic glass – unusual properties from glassy nano-domains. *Phys. Stat. Solidi B* **251**, 1982–1992 (2014).
8. Yeh, J.-W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
9. Gludovatz, B. et al. A fracture-resistant high-entropy alloy for cryogenic applications. *Science* **345**, 1153–1158 (2014).
10. Li, Z., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength-ductility trade-off. *Nature* **534**, 227–230 (2016).
11. Lei, Z. et al. Enhanced strength and ductility in a high-entropy alloy via ordered oxygen complexes. *Nature* **563**, 546–550 (2018); correction **565**, E8 (2019).
12. Yang, T. et al. Multicomponent intermetallic nanoparticles and superb mechanical behaviors of complex alloys. *Science* **362**, 933–937 (2018).
13. Chung, D., Ding, Z. & Yang, Y. Hierarchical eutectic structure enabling superior fracture toughness and superb strength in

CoCrFeNiNb0.5 eutectic high entropy alloy at room temperature. *Adv. Eng. Mater.* **21**, 1801060 (2019).

14. Hua, N. et al. Mechanical, corrosion, and wear properties of biomedical Ti–Zr–Nb–Ta–Mo high entropy alloys. *J. Alloys Compd.* **861**, 157997 (2021).
15. Li, Y. et al. Research progress on refractory high entropy alloys. *Rare Met. Mater. Eng.* **49**, 4365–4372 (2020).
16. Ye, Y. F., Wang, Q., Lu, J., Liu, C. T. & Yang, Y. High-entropy alloy: challenges and prospects. *Mater. Today* **19**, 349–362 (2016).
17. Ye, Y. F., Liu, C. T. & Yang, Y. A geometric model for intrinsic residual strain and phase stability in high entropy alloys. *Acta Mater.* **94**, 152–161 (2015).
18. Ye, Y. F. et al. Atomic-scale distorted lattice in chemically disordered equimolar complex alloys. *Acta Mater.* **150**, 182–194 (2018).
19. Larsen, P. M., Schmidt, S. & Schiøtz, J. Robust structural identification via polyhedral template matching. *Modell. Simul. Mater. Sci. Eng.* **24**, 055007 (2016).
20. Pugno, N. M. & Ruoff, R. S. Nanoscale Weibull statistics. *J. Appl. Phys.* **99**, 024301 (2006).
21. Yang, Y., Ye, J. C., Lu, J., Liu, F. X. & Liaw, P. K. Effects of specimen geometry and base material on the mechanical behavior of focused-ion-beam-fabricated metallic-glass micropillars. *Acta Mater.* **57**, 1613–1623 (2009).
22. Greer, J. R. & Nix, W. D. Nanoscale gold pillars strengthened through dislocation starvation. *Phys. Rev. B* **73**, 245410 (2006).
23. Kiener, D., Motz, C., Schöberl, T., Jenko, M. & Dehm, G. Determination of mechanical properties of copper at the micron scale. *Adv. Eng. Mater.* **8**, 1119–1125 (2006).

24. Dimiduk, D. M., Uchic, M. D. & Parthasarathy, T. A. Size-affected single-slip behavior of pure nickel microcrystals. *Acta Mater.* **53**, 4065–4077 (2005).
25. Frick, C. P., Clark, B. G., Orso, S., Schneider, A. S. & Arzt, E. Size effect on strength and strain hardening of small-scale [111] nickel compression pillars. *Mater. Sci. Eng. A* **489**, 319–329 (2008).
26. Zou, Y., Maiti, S., Steurer, W. & Spolenak, R. Size-dependent plasticity in an Nb₂₅Mo₂₅Ta₂₅W₂₅ refractory high-entropy alloy. *Acta Mater.* **65**, 85–97 (2014).
27. Chen, Z. M. T., Okamoto, N. L., Demura, M. & Inui, H. Micropillar compression deformation of single crystals of Co₃(Al,W) with the L1₂ structure. *Scripta Mater.* **121**, 28–31 (2016).
28. Gómez-Cortés, J. F. et al. Size effect and scaling power-law for superelasticity in shape-memory alloys at the nanoscale. *Nat. Nanotechnol.* **12**, 790–796 (2017).
29. Li, F. C. et al. The stochastic transition from size dependent to size independent yield strength in metallic glasses. *J. Mech. Phys. Solids* **109**, 200–216 (2017).
30. Lai, Y. H. et al. Bulk and microscale compressive behavior of a Zr-based metallic glass. *Scripta Mater.* **58**, 890–893 (2008).
31. Ye, J. C., Lu, J., Yang, Y. & Liaw, P. K. Extraction of bulk metallic-glass yield strengths using tapered micropillars in micro-compression experiments. *Intermetallics* **18**, 385–393 (2010).
32. Ye, J. C., Lu, J., Liu, C. T., Wang, Q. & Yang, Y. Atomistic free-volume zones and inelastic deformation of metallic glasses. *Nat. Mater.* **9**, 619–623 (2010).
33. Hao, S. et al. Achieving large linear elasticity and high strength in bulk nanocomposite via synergistic effect. *Sci. Rep.* **5**, 8892 (2015).

34. Zhang, L., Xiang, Y., Han, J. & Srolovitz, D. J. The effect of randomness on the strength of high-entropy alloys. *Acta Mater.* **166**, 424–434 (2019).
35. Han, S. M. et al. Critical-temperature/Peierls-stress dependent size effects in body centered cubic nanopillars. *Appl. Phys. Lett.* **102**, 041910 (2013).
36. Lee, S.-W. & Nix, W. D. Size dependence of the yield strength of fcc and bcc metallic micropillars with diameters of a few micrometers. *Philos. Mag.* **92**, 1238–1260 (2012).
37. Williams, D. B. & Carter, C. B. *Transmission Electron Microscopy: A Textbook for Materials Science* 271–282 (Springer, 2009).
38. Feuerbacher, M. Dislocations and deformation microstructure in a B2-ordered Al₂₈Co₂₀Cr₁₁Fe₁₅Ni₂₆ high-entropy alloy. *Sci. Rep.* **6**, 29700 (2016).
39. Gschneidner, K. Jr et al. A family of ductile intermetallic compounds. *Nat. Mater.* **2**, 587–591 (2003).
40. Wu, Z., Gao, Y. & Bei, H. Thermal activation mechanisms and Labusch-type strengthening analysis for a family of high-entropy and equiatomic solid-solution alloys. *Acta Mater.* **120**, 108–119 (2016).
41. Kamimura, Y., Edagawa, K. & Takeuchi, S. Experimental evaluation of the Peierls stresses in a variety of crystals and their relation to the crystal structure. *Acta Mater.* **61**, 294–309 (2013).
42. Laplanche, G. et al. Elastic moduli and thermal expansion coefficients of medium-entropy subsystems of the CrMnFeCoNi high-entropy alloy. *J. Alloys Compd.* **746**, 244–255 (2018).
43. Hausch, G. Elastic and magnetoelastic effects in invar alloys. *J. Magn. Magn. Mater.* **10**, 163–169 (1979).

44. Wasserman, E. F. In *Handbook of Ferromagnetic Materials* Vol. 5 (eds Buschow, K. H. J. & Wohlfarth, E. P.) 237–322 (Elsevier, 1990).
45. Lam, N. Q. & Okamoto, P. R. A unified approach to solid-state amorphization and melting. *MRS Bull.* **19**, 41–46 (1994).
46. Wang, W. H., Bai, H. Y., Luo, J. L., Wang, R. J. & Jin, D. Supersoftening of transverse phonons in $Zr_{41}Ti_{14}Cu_{12.5}Ni_{10}B_{22.5}$ bulk metallic glass. *Phys. Rev. B* **62**, 25–28 (2000).
47. Schuh, C. A., Hufnagel, T. C. & Ramamurty, U. Mechanical behavior of amorphous alloys. *Acta Mater.* **55**, 4067–4109 (2007).
48. Grover, R., Getting, I. C. & Kennedy, G. C. Simple compressibility relation for solids. *Phys. Rev. B* **7**, 567–571 (1973).
49. Liang, L., Ma, H. & Wei, Y. Size-dependent elastic modulus and vibration frequency of nanocrystals. *J. Nanomater.* **2011**, 670857 (2011).
50. Kalidindi, S. R., Abusafieh, A. & El-Danaf, E. Accurate characterization of machine compliance for simple compression testing. *Exp. Mech.* **37**, 210–215 (1997).
51. Wang, W. H. The elastic properties, elastic models and elastic perspectives of metallic glasses. *Prog. Mater Sci.* **57**, 487–656 (2012).
52. Etienne, S., Cavaille, J. Y., Perez, J., Point, R. & Salvia, M. Automatic system for analysis of micromechanical properties. *Rev. Sci. Instrum.* **53**, 1261–1266 (1982).
53. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
54. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B* **48**, 13115–13118 (1993).

55. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
56. Perdew, J. P. et al. Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992); erratum **48**, 4978 (1993).
57. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **45**, 13244–13249 (1992).
58. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
59. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
60. von Mises, R. Mechanics of solid bodies in the plastically-deformable state. *Göttin. Nachr. Math. Phys.* **1**, 582–592 (1913).
61. Shimizu, F., Ogata, S. & Li, J. Theory of shear banding in metallic glasses and molecular dynamics calculations. *Mater. Trans.* **48**, 2923–2927 (2007).
62. Wang, S. Q. & Ye, H. Q. Ab initioelastic constants for the lonsdaleite phases of C, Si and Ge. *J. Phys. Condens. Matter* **15**, 5307–5314 (2003).
63. Grimvall, G. *Thermophysical Properties of Materials* 27–45 (North Holland, 1999).
64. Zhang, J.-M., Zhang, Y., Xu, K.-W. & Ji, V. Representation surfaces of Young’s modulus and Poisson’s ratio for BCC transition metals. *Physica B* **390**, 106–111 (2007).

65. Parrinello, M. & Rahman, A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).
66. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
67. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* 2nd edn (Oxford Univ. Press, 2017).

Acknowledgements

D.J.S. gratefully acknowledges the support of the Research Grant Council, Hong Kong Government, through the General Research Fund (GRF) with grant nos HKU 11211019. J.G.W. acknowledges the support of Guangdong Major Project of Basic and Applied Basic Research, China (grant no. 2019B030302010). The research of Y.Y. is supported by the Research Grant Council, Hong Kong Government, through the General Research Fund (GRF) with the grant nos CityU11213118 and CityU11200719 as well as by the City University of Hong Kong with grant no. 9610391. C.W.P. acknowledges the financial support of Academia Sinica Career Development Award with grant no. 2317-1050100. Q.Z. acknowledges the funding from the National Natural Science Foundation of China (Nos. 51871054). C.W.P. is grateful for computational support from the National Center for High-performance Computing, Taiwan. Q.F.H. is grateful for the assistance given by X. K. Xi, X. D. Liu and T. Liu.

Author information

Author notes

1. These authors contributed equally: Q. F. He, J. G. Wang, H. A. Chen

Affiliations

1. Department of Mechanical Engineering, City University of Hong Kong, Kowloon, Hong Kong, China

Q. F. He, J. G. Wang, Z. Y. Ding, Z. Q. Zhou, J. C. Qiao, C. T. Liu & Y. Yang
2. College of Mechanical Engineering, Dongguan University of Technology, Dongguan, China

J. G. Wang
3. Research Center for Applied Sciences, Academia Sinica, Taipei, Taiwan

H. A. Chen & C. W. Pao
4. X-ray Science Division, Argonne National Laboratory, Lemont, IL, USA

L. H. Xiong & Y. Ren
5. Department of Materials Science and Engineering, Hong Kong Institute for Advanced Study, City University of Hong Kong, Kowloon, Hong Kong, China

J. H. Luan, C. T. Liu & Y. Yang
6. MATEIS, Université de Lyon, UMR CNRS5510, INSA-Lyon, Villeurbanne, France

J. M. Pelletier & J. C. Qiao
7. School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an, China

J. C. Qiao
8. Laboratory for Structures, Institute of Materials, Shanghai University, Shanghai, China

Q. Wang

9. College of Physics and Materials Science, Tianjin Normal University, Tianjin, China

L. L. Fan

10. Department of Physics, City University of Hong Kong, Kowloon, Hong Kong, China

Y. Ren

11. Center for High Pressure Science and Technology Advanced Research, Shanghai, China

Q. S. Zeng

12. Jiangsu Key Laboratory of Advanced Metallic Materials, School of Materials Science and Engineering, Southeast University, Nanjing, China

Q. S. Zeng

13. Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China

D. J. Srolovitz

14. International Digital Economy Academy (IDEA), Shenzhen, China

D. J. Srolovitz

Contributions

Y.Y. supervised the project. Y.Y., C.W.P. and D.J.S. conceived the idea. Q.F.H. fabricated the polycrystalline samples and J.C.Q. prepared the single-crystal samples. Q.F.H. characterized the structures and mechanical properties of the samples. J.G.W., C.W.P. and H.A.C. carried out the atomistic simulations. J.C.Q. and J.M.P. performed the dynamic mechanical

spectroscopy analyses. J.H.L. and C.T.L. performed the 3D APT experiments. L.H.X., L.L.F., Q.S.Z. and Y.R. performed the in situ HEXRD experiments. Y.Y., D.J.S., C.W.P., Z.Y.D., Z.Q.Z. and Q.W. contributed to the data analysis. Y.Y., Q.F.H. and C.W.P. wrote the manuscript. All authors participated in the discussion of the results.

Corresponding authors

Correspondence to [C. W. Pao](#), [D. J. Srolovitz](#) or [Y. Yang](#).

Ethics declarations

Competing interests

Y.Y. and Q.F.H are in the process of applying a patent related to the alloy design described in this work. The remaining authors declare no competing interests.

Peer review information

Nature thanks David Dye, Andrew Minor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 Characterization of single-crystal Co₂₅Ni₂₅\(HfTiZr\)₅₀ alloy.](#)

a, A low-magnification backscatter electron image and elemental distributions show that the chemical distribution is homogeneous in the

single-crystal samples on a sub-micro scale. **b**, Low-magnification TEM image and the corresponding diffraction patterns ([001] zone axis) in different regions show that there is no phase separation in the single-crystal samples.

Extended Data Fig. 2 The characterization of the polycrystalline Co₂₅Ni₂₅(HfTiZr)₅₀.

a, The inverse pole figure map showing the grain structure of as-cast polycrystalline Co₂₅Ni₂₅(HfTiZr)₅₀ alloy. **b**, The pole figures showing that there are no preferred orientations in the as-cast polycrystalline Co₂₅Ni₂₅(HfTiZr)₅₀ alloy. **c**, The XRD patterns of the Co₂₅Ni₂₅(HfTiZr)₅₀ alloy samples after thermal annealing at 1,273 K for different time durations all exhibit single-phase B2 ordering. **d**, Compression stress–strain curves of the single-crystal and polycrystal Co₂₅Ni₂₅(HfTiZr)₅₀ alloy after annealing at 1,273 K for 9 h. The results show that the mechanical properties of the Co₂₅Ni₂₅(HfTiZr)₅₀ alloy do not change after the heat treatment. **e**, A low-magnification SEM image shows the microstructure of the Co₂₅Ni₂₅(HfTiZr)₅₀ alloy after annealing at 1,273 K for 9 h. **f**, The STEM image and elemental mapping near a grain boundary. No segregation to grain boundaries was observed following a 9-h, 1,273-K anneal in our Co₂₅Ni₂₅(HfTiZr)₅₀ alloy. (AC and HT represent as-cast and heat-treated, respectively.) **g**, The APT reconstructions of the three-dimensional elemental distributions showing the chemical homogeneity along the grain boundary at the nanometre scale. The black rectangle in the EBSD image indicates the position from which the APT tip was carved.

Extended Data Fig. 3 The monotonic and cyclic microcompression results of the single-crystalline Co₂₅Ni₂₅(HfTiZr)₅₀.

a, The typical monotonic stress–strain curves obtained for different micropillar diameters. The inset shows a typical pillar image. **b**, The Young’s modulus versus pillar diameter of the micropillars in Fig. 3b. The average Young’s modulus is measured to be 106 GPa. **c**, The size-dependent

yield strength of single-crystal $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ micropillars. **d**, The cyclic stress-strain curves obtained within the elastic regime from the micropillar at different nominal stress rates. Note that no mechanical hysteresis is evident in **d**. The inset shows the (cyclic) load versus time. Five cycles were performed for each compression.

Extended Data Fig. 4 Loss factors.

The loss factors measured for the single-crystal and polycrystalline $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ in comparison with various bulk metallic glasses over a wide temperature range.

Extended Data Fig. 5 Comparison of the first height of the steel ball bouncing back from different alloy surfaces.

a, Photo showing the starting moment of the bouncing experiments. **b–e**, The first height of the steel ball after bouncing back, as indicated by the white arrow, from single-crystalline $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ (**b**), spark plasma sintered (SPS) $\text{Cu}_{50}\text{Zr}_{45}\text{Al}_5$ metallic glass (**c**), NiAl alloy with a B2 structure (**d**), and commercial stainless steel (**e**). See Supplementary Video [1](#) for details. We note that all the bulk alloys had a similar size.

Extended Data Fig. 6 The dislocation structure analysis performed in the [111] single-crystalline $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ sample after deforming to 4% mechanical strain.

a, $(g=(0\bar{2}0))$ and beam direction $Z=[001]$. **b**, $(g=(\bar{1}\bar{1}0))$ and beam direction $Z=[001]$. **c**, $g=(200)$ and beam direction $Z=[001]$. **d**, $(g=(0\bar{1}1))$ and beam direction $(Z=[\bar{1}11])$. **e**, $(g=(\bar{1}\bar{1}0))$ and beam direction $(Z=[\bar{1}11])$. **f**, $g=(101)$ and beam direction $(Z=[\bar{1}11])$. The $\mathbf{g} \cdot \mathbf{b} = 0$ out of contrast analyses indicate that the dislocations are of the $\langle 001 \rangle$ type. See Supplementary Table [1](#) for detailed analysis and description of labels A–D.

Extended Data Fig. 7 Stress relaxation and internal friction stress.

a, The typical stress relaxation curve obtained from a single-crystalline $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ micropillar with a top diameter of $1\ \mu\text{m}$. The activation volume is calculated to be $\sim 3.05b^3$. **b**, The yield strength of single-crystal $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ micropillars at different temperatures. Standard fits to the data (inset equation) yield a Peierls stress of $\tau_c \approx 0.47\sigma_c = 2.8\ \text{GPa}$ and effective temperature of $T_0 = 1,107\ \text{K}$. The inset shows the contour plot of the critical stress τ_c as a function of the correlation length λ and standard deviation Δ for $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$. Note that ζ_0 stands for a dislocation core size.

Extended Data Fig. 8 The magnetic properties measured for the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.

a, The magnetization curve of the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy as a function of the applied magnetic field at room temperature. The saturation magnetization M_s of $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ is only $1.17\ \text{emu g}^{-1}$. **b**, The temperature dependence of magnetization of $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ under the applied magnetic field of 500 Oe. The result shows that there is an antiferromagnetic (AFM) to ferromagnetic (FM) transition at the transition temperature $T_N = 851\ \text{K}$. **c**, The measured magnetostriction coefficient along different directions of single-crystal $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy. The magnetostriction coefficient of $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy is about zero. The ferromagnetic polycrystalline Ni- and Fe-based metallic glass (MG) are taken for comparison.

Extended Data Fig. 9 The linear thermal expansion coefficient of the $\text{Co}_{25}\text{Ni}_{25}(\text{HfTiZr})_{50}$ alloy.

a, The thermal expansion curves obtained from experiments. The average thermal expansion coefficients (α) of the single-crystal and polycrystalline

samples are almost the same, about $11.4 \times 10^{-6} \text{ K}^{-1}$. **b**, The variation of the lattice constant with temperature calculated from the ab initio molecular dynamics simulations. The average thermal expansion coefficient is about $8.1 \times 10^{-6} \text{ K}^{-1}$, which is very close to our experimental measurement.

Supplementary information

Supplementary Information

This file contains Supplementary Text, Supplementary Equations, Supplementary Figure 1, Supplementary Table 1, Supplementary References and the legend for Supplementary Video 1

Supplementary Video 1

Demonstration of elasticity of different metals with steel ball bouncing tests.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

He, Q.F., Wang, J.G., Chen, H.A. *et al.* A highly distorted ultralastic chemically complex Elinvar alloy. *Nature* **602**, 251–257 (2022).
<https://doi.org/10.1038/s41586-021-04309-1>

- Received: 01 March 2021
- Accepted: 06 December 2021
- Published: 09 February 2022

- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04309-1>

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/s41586-021-04309-1>

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

- Article
- [Published: 09 February 2022](#)

Superionic iron alloys and their seismic velocities in Earth's inner core

- [Yu He](#) [ORCID: orcid.org/0000-0001-6518-585X](#)^{1,2 na1},
- [Shichuan Sun](#)^{1,3 na1},
- [Duck Young Kim](#) [ORCID: orcid.org/0000-0002-0765-6374](#)²,
- [Bo Gyu Jang](#)²,
- [Heping Li](#)¹ &
- [Ho-kwang Mao](#)²

[Nature](#) volume **602**, pages 258–262 (2022)

- 1334 Accesses
- 365 Altmetric
- [Metrics details](#)

Subjects

- [Geophysics](#)
- [Mineralogy](#)
- [Phase transitions and critical phenomena](#)

Abstract

Earth's inner core (IC) is less dense than pure iron, indicating the existence of light elements within it¹. Silicon, sulfur, carbon, oxygen and hydrogen have been suggested to be the candidates^{2,3}, and the properties of iron–light-element alloys have been studied to constrain the IC composition^{4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19}. Light elements have a substantial influence on the seismic velocities^{4,5,6,7,8,9,10,11,12,13}, the melting temperatures^{14,15,16,17} and the thermal conductivities^{18,19} of iron alloys. However, the state of the light elements in the IC is rarely considered. Here, using ab initio molecular dynamics simulations, we find that hydrogen, oxygen and carbon in hexagonal close-packed iron transform to a superionic state under the IC conditions, showing high diffusion coefficients like a liquid. This suggests that the IC can be in a superionic state rather than a normal solid state. The liquid-like light elements lead to a substantial reduction in the seismic velocities, which approach the seismological observations of the IC^{20,21}. The substantial decrease in shear-wave velocity provides an explanation for the soft IC²¹. In addition, the light-element convection has a potential influence on the IC seismological structure and magnetic field.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Phase diagrams of $\text{FeH}_{0.25}$, $\text{FeO}_{0.0625}$ and $\text{FeC}_{0.0625}$ at temperatures and pressures in the range of about 2,000–7,000 K and about 250–400 GPa.



Fig. 2: Diffusion coefficients of light elements (H, O and C) in solid and liquid Fe alloys under the IC conditions.



Fig. 3: Seismic velocities of $\text{FeH}_{0.25}$, $\text{FeO}_{0.0625}$ and $\text{FeC}_{0.0625}$ as a function of density and temperature at 360 GPa.



Fig. 4: Schematic of the outer-core fluid convection and the IC light-element convection.



Data availability

The data supporting the findings of this study have been deposited at the 4TU Center for Research Data: <https://doi.org/10.4121/12932588.v2>. Any additional data are available upon request from the corresponding author. [Source data](#) are provided with this paper.

Code availability

The Vienna Ab initio Simulation Package is a proprietary software available for purchase at <https://www.vasp.at/>. Phonopy code is available at <http://phonopy.github.io/phonopy/>. WIEN2k is available at <http://www.wien2k.at/>. The WIEN2k+eDMFT package is available at <http://hauleweb.rutgers.edu/tutorials/>.

References

1. Birch, F. Elasticity and constitution of the Earth's interior. *J. Geophys. Res.* **57**, 227–286 (1952).
2. Poirier, J. P. & Shankland, T. J. Dislocation melting of iron and temperature of the inner core boundary, revisited. *Geophys. J. Int.* **115**, 147–151 (1993).
3. Li, J. & Fei, Y. in *Treatise on Geochemistry, Vol. 2: The Mantle and Core* (ed. Carlson, R. W.) 1–31 (Elsevier, 2007).
4. Li, J., Fei, Y., Mao, H. K., Hirose, K. & Shieh, S. R. Sulfur in the Earth's inner core. *Earth Planet. Sci. Lett.* **193**, 509–514 (2001).
5. Badro, J. et al. Effect of light elements on the sound velocity of solid iron: implications for the composition of Earth's core. *Earth Planet. Sci. Lett.* **254**, 233–238 (2007).
6. Mao, Z. et al. Sound velocities of Fe and Fe–Si alloy in the Earth's core. *Proc. Natl Acad. Sci. USA* **109**, 10239–10244 (2012).
7. Lin, J. F., Heinz, D. L., Campbell, A. J., Devine, J. M. & Shen, G. Y. Iron–silicon alloy in Earth's core? *Science* **295**, 313–315 (2002).

8. Lin, J. F. et al. Sound velocities of iron–nickel and iron–silicon alloys at high pressures. *Geophys. Res. Lett.* **30**, 2112 (2003).
9. Antonangeli, D. et al. Composition of the Earth’s inner core from high-pressure sound velocity measurements in Fe–Ni–Si alloys. *Earth Planet. Sci. Lett.* **295**, 292–296 (2010).
10. Chen, B. et al. Hidden carbon in Earth’s inner core revealed by shear softening in dense Fe_7C_3 . *Proc. Natl Acad. Sci. USA* **111**, 17755–17758 (2014).
11. Prescher, C. et al. High Poisson’s ratio of Earth’s inner core explained by carbon alloying. *Nat. Geosci.* **8**, 220–223 (2015).
12. Shibasaki, Y. et al. Sound velocity measurements in dhcp-FeH up to 70 GPa with inelastic X-ray scattering: implications for the composition of the Earth’s core. *Earth Planet. Sci. Lett.* **313–314**, 79–85 (2012).
13. Caracas, R. The influence of hydrogen on the seismic properties of solid iron. *Geophys. Res. Lett.* **42**, 3780–3785 (2015).
14. Alfè, D., Gillan, M. J. & Price, G. D. Composition and temperature of the Earth’s core constrained by combining ab initio calculations and seismic data. *Earth Planet. Sci. Lett.* **195**, 91–98 (2002).
15. Terasaki, H. et al. Liquidus and solidus temperatures of a Fe–O–S alloy up to the pressures of the outer core: implication for the thermal structure of the Earth’s core. *Earth Planet. Sci. Lett.* **304**, 559–564 (2011).
16. Morard, G. et al. Fe–FeO and Fe– Fe_3C melting relations at Earth’s core–mantle boundary conditions: implications for a volatile-rich or oxygen-rich core. *Earth Planet. Sci. Lett.* **473**, 94–103 (2017).
17. Mashino, I., Miozzi, F., Hirose, K., Morard, G. & Sinmyo, R. Melting experiments on the Fe–C binary system up to 255 GPa: constraints on the carbon content in the Earth’s core. *Earth Planet. Sci. Lett.* **515**, 135–144 (2019).

18. Pozzo, M., Davies, C., Gubbins, D. & Alfè, D. Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* **485**, 355–358 (2012).
19. de Koker, N., Steinle-Neumann, G. & Vlcek, V. Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T , and heat flux in Earth's core. *Proc. Natl Acad. Sci. USA* **109**, 4070–4073 (2012).
20. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
21. Tkalčić, H. & Pham, T.-S. Shear properties of Earth's inner core constrained by a detection of J waves in global correlation wavefield. *Science* **362**, 329–332 (2018).
22. Martorell, B., Vočadlo, L., Brodholt, J. & Wood, I. G. Strong premelting effect in the elastic properties of hcp-Fe under inner-core conditions. *Science* **342**, 466–468 (2013).
23. Martorell, B., Wood, I. G., Brodholt, J. & Vočadlo, L. The elastic properties of hcp- $\text{Fe}_{1-x}\text{Si}_x$ at Earth's inner-core conditions. *Earth Planet. Sci. Lett.* **451**, 89–96 (2016).
24. Li, Y., Vočadlo, L. & Brodholt, J. B. The elastic properties of hcp-Fe alloys under the conditions of the Earth's inner core. *Earth Planet. Sci. Lett.* **493**, 118–127 (2018).
25. Singh, S. C., Taylor, M. A. J. & Montagner, J. P. On the presence of liquid in Earth's inner core. *Science* **287**, 2471–2474 (2000).
26. Creager, K. C. Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP. *Nature* **356**, 309–314 (1992).
27. Vočadlo, L., Dobson, D. P. & Wood, I. G. Ab initio calculations of the elasticity of hcp-Fe as a function of temperature at inner-core pressure. *Earth Planet. Sci. Lett.* **288**, 534–538 (2009).

28. Belonoshko, A. B., Skorodumova, N. V., Rosengren, A. & Johansson, B. Elastic anisotropy of Earth's inner core. *Science* **319**, 797–800 (2008).
29. Song, X. & Richards, P. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **382**, 221–224 (1996).
30. Su, W. J., Dziewonski, A. M. & Jeanloz, R. Planet within a planet: rotation of the inner core of Earth. *Science* **274**, 1883–1887 (1996).
31. Tkalčić, H., Young, M., Bodin, T., Ngo, S. & Sambridge, M. The shuffling rotation of the Earth's inner core revealed by earthquake doublets. *Nat. Geosci.* **6**, 497–502 (2013).
32. Lodders, K. Solar System abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247 (2003).
33. Li, Y., Vočadlo, L., Alfè, D. & Brodholt, J. B. Carbon partitioning between the Earth's inner and outer core. *J. Geophys. Res. Solid Earth* **124**, 12812–12824 (2019).
34. Pamato, M. G. et al. Equation of state of hcp Fe–C–Si alloys and the effect of C incorporation mechanism on the density of hcp Fe alloys at 300 K. *J. Geophys. Res. Solid Earth* **125**, e2020JB020159 (2020).
35. Alfè, D. Temperature of the inner-core boundary of the Earth: melting of iron at high pressure from first-principles coexistence simulations. *Phys. Rev. B* **79**, 060101 (2009).
36. Anzellini, S., Dewaele, A., Mezouar, M., Loubeyre, P. & Morard, G. Melting of iron at Earth's inner core boundary based on fast X-ray diffraction. *Science* **340**, 464–466 (2013).
37. Cavazzoni, C. et al. Superionic and metallic states of water and ammonia at giant planet conditions. *Science* **283**, 44–46 (1999).
38. Goldman, N., Fried, L. E., Kuo, I.-F. W. & Mundy, C. J. Bonding in the superionic phase of water. *Phys. Rev. Lett.* **94**, 217801 (2005).

39. Hernandez, J.-A. & Caracas, R. Superionic–superionic phase transitions in body-centered cubic H₂O ice. *Phys. Rev. Lett.* **117**, 135503 (2016).
40. Millot, M. et al. Nanosecond X-ray diffraction of shock-compressed superionic water ice. *Nature* **569**, 251–255 (2019).
41. Millot, M. et al. Experimental evidence for superionic water ice using shock compression. *Nat. Phys.* **14**, 297–302 (2018).
42. Hou, M. et al. Superionic iron oxide–hydroxide in Earth’s deep mantle. *Nat. Geosci.* **14**, 174–178 (2021).
43. Hull, S., Farley, T. W. D., Hayes, W. & Hutchings, M. T. The elastic properties of lithium oxide and their variation with temperature. *J. Nucl. Mater.* **160**, 125–134 (1988).
44. Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth’s inner core. *Science* **274**, 1887–1891 (1996).
45. Tateno, S., Hirose, K., Ohishi, Y. & Tatsumi, Y. The structure of iron in Earth’s inner core. *Science* **330**, 359–361 (2010).
46. Belonoshko, A. B., Rajeev, A. & Johansson, B. Stability of the body-centred-cubic phase of iron in the Earth’s inner core. *Nature* **424**, 1032–1034 (2003).
47. Sakamaki, T. et al. Constraints on Earth’s inner core composition inferred from measurements of the sound velocity of hcp-iron in extreme conditions. *Sci. Adv.* **2**, e1500802 (2016).
48. Antonangeli, D. et al. Sound velocities and density measurements of solid hcp-Fe and hcp-Fe–Si (9 wt.%) alloy at high pressure: constraints on the Si abundance in the Earth’s inner core. *Earth Planet. Sci. Lett.* **482**, 446–453 (2018).
49. Caracas, R. The influence of carbon on the seismic properties of solid iron. *Geophys. Res. Lett.* **44**, 128–134 (2017).

50. Tagawa, S., Ohta, K., Hirose, K., Kato, C. & Ohishi, Y. Compression of Fe–Si–H alloys to core pressures. *Geophys. Res. Lett.* **43**, 3686–3692 (2016).
51. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
52. Klarbring, J. & Simak, S. I. Phase stability of dynamically disordered solids from first principles. *Phys. Rev. Lett.* **121**, 225702 (2018).
53. Terasaki, H. et al. Stability of Fe–Ni hydride after the reaction between Fe–Ni alloy and hydrous phase (δ -AlOOH) up to 1.2 Mbar: possibility of H contribution to the core density deficit. *Phys. Earth Planet. Inter.* **194–195**, 18–24 (2012).
54. Ozawa, H., Hirose, K., Tateno, S., Sata, N. & Ohishi, Y. Phase transition boundary between B1 and B8 structures of FeO up to 210 GPa. *Phys. Earth Planet. Inter.* **179**, 157–163 (2010).
55. Tateno, S., Kuwayama, Y., Hirose, K. & Ohishi, Y. The structure of Fe–Si alloy in Earth’s inner core. *Earth Planet. Sci. Lett.* **418**, 11–19 (2015).
56. Mori, Y. et al. Melting experiments on Fe–Fe₃S system to 254 GPa. *Earth Planet. Sci. Lett.* **464**, 135–141 (2017).
57. Fischer, R. A. et al. Equation of state and phase diagram of Fe–16Si alloy as a candidate component of Earth’s core. *Earth Planet. Sci. Lett.* **357–358**, 268–276 (2012).
58. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
59. Kresse, G. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

60. Alfè, D. & Gillan, M. J. First-principles calculation of transport coefficients. *Phys. Rev. Lett.* **81**, 5161–5164 (1998).
61. He, Y. et al. First-principles prediction of fast migration channels of potassium ions in KAlSi_3O_8 hollandite: implications for high conductivity anomalies in subduction zones. *Geophys. Res. Lett.* **43**, 6228–6233 (2016).
62. Mookherjee, M., Stixrude, L. & Karki, B. Hydrous silicate melt at high pressure. *Nature* **452**, 983–986 (2008).
63. Belonoshko, A. B. et al. Stabilization of body-centred cubic iron under inner-core conditions. *Nat. Geosci.* **10**, 312–316 (2017).
64. Alfè, D., Cazorla, C. & Gillan, M. J. The kinetics of homogeneous melting beyond the limit of superheating. *J. Chem. Phys.* **135**, 024102 (2011).
65. Alfè, D., Gillan, M. J. & Price, G. D. Temperature and composition of the Earth's core. *Contemp. Phys.* **48**, 63–80 (2007).
66. Alfè, D. First-principles simulations of direct coexistence of solid and liquid aluminum. *Phys. Rev. B* **68**, 064423 (2003).
67. Bonev, S. A., Schwegler, E., Ogitsu, T. & Galli, G. A quantum fluid of metallic hydrogen suggested by first-principles calculations. *Nature* **431**, 669–672 (2004).
68. Ogitsu, T., Schwegler, E., Gygi, F. & Galli, G. Melting of lithium hydride under pressure. *Phys. Rev. Lett.* **91**, 175502 (2003).
69. Alfè, D. Melting curve of MgO from first-principles simulations. *Phys. Rev. Lett.* **94**, 235701 (2005).
70. Yoo, C. S., Holmes, N. C., Ross, M., Webb, D. J. & Pike, C. Shock temperatures and melting of iron at Earth core conditions. *Phys. Rev. Lett.* **70**, 3931–3934 (1993).

71. Alfè, D., Price, G. D. & Gillan, M. J. Iron under Earth's core conditions: liquid-state thermodynamics and high-pressure melting curve from ab initio calculations. *Phys. Rev. B* **65**, 165118 (2002).
72. Sola, E. & Alfè, D. Melting of iron under Earth's core conditions from diffusion Monte Carlo free energy calculations. *Phys. Rev. Lett.* **103**, 078501 (2009).
73. Haule, K., Yee, C.-H. & Kim, K. Dynamical mean-field theory within the full-potential methods: electronic structure of CeIrIn₅, CeCoIn₅, and CeRhIn₅. *Phys. Rev. B* **81**, 195107 (2010).
74. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **78**, 1396 (1997).
75. Xu, J. et al. Thermal conductivity and electrical resistivity of solid iron at Earth's core conditions from first principles. *Phys. Rev. Lett.* **121**, 096601 (2018).
76. Gomi, H. et al. The high conductivity of iron and thermal evolution of the Earth's core. *Phys. Earth Planet. Inter.* **224**, 88–103 (2013).
77. Sun, S., He, Y., Kim, D. Y. & Li, H. Anomalous elastic properties of superionic ice. *Phys. Rev. B* **102**, 104108 (2020).
78. He, Y., Sun, S. & Li, H. Ab initio molecular dynamics investigation of the elastic properties of superionic Li₂O under high temperature and pressure. *Phys. Rev. B* **103**, 174105 (2021).
79. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).

Acknowledgements

This study was supported by the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (XDB 18010401). We acknowledge the support of the National Natural Science Foundation of China

(42074104, 41774101, 11774015, U1930401) and the Youth Innovation Promotion Association of CAS (2020394). Numerical computations were performed at the Hefei Advanced Computing Center, the Shanghai Supercomputer Center and the National Supercomputer Center in Guangzhou.

Author information

Author notes

1. These authors contributed equally: Yu He, Shichuan Sun

Affiliations

1. Key Laboratory of High-Temperature and High-Pressure Study of the Earth's Interior, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, China

Yu He, Shichuan Sun & Heping Li

2. Center for High Pressure Science and Technology Advanced Research, Shanghai, China

Yu He, Duck Young Kim, Bo Gyu Jang & Ho-kwang Mao

3. University of Chinese Academy of Sciences, Beijing, China

Shichuan Sun

Contributions

Y.H. and S.S. contributed equally to this work, conducted the calculations, analysed the data and wrote the manuscript. Y.H., D.Y.K. and H.-k.M. initiated and designed the project. Y.H. conducted calculations on phase transition, melting temperatures and diffusion properties. S.S. performed simulations on elastic properties. B.G.J. performed electronic conductivity calculations. Y.H., D.Y.K. and H.L. discussed the geophysical implications.

All authors discussed the data interpretation and commented on the manuscript.

Corresponding author

Correspondence to [Yu He](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Trajectories and mean square displacements \(MSDs\) of H, O, C and Fe in Fe-light-element alloys.](#)

Trajectories of **a**, H and Fe in $\text{FeH}_{0.25}$; **b**, O and Fe in $\text{FeO}_{0.0625}$; **c**, C and Fe in $\text{FeC}_{0.0625}$ at the superionic state under the IC conditions (~ 360 GPa and ~ 5000 K). Small pink, red, black and brown spheres represent the

trajectories of H, O, C and Fe, respectively. MSDs of **d**, H and Fe in $\text{FeH}_{0.25}$; **e**, O and Fe in $\text{FeO}_{0.0625}$; **f**, C and Fe in $\text{FeC}_{0.0625}$.

[Source data](#)

Extended Data Fig. 2 The evolution of temperature and pressure with respect to simulation time in the two-phase coexisting systems for Fe-H, Fe-O and Fe-C alloys.

a, Temperatures and **b**, pressures for Fe-H, Fe-O and Fe-C alloys are shown with light grey, pink and cyan curves, and the averaged data over a 0.5 ps period are shown with thick black, red, and blue curves.

[Source data](#)

Extended Data Fig. 3 Ionic conductivities of superionic Fe alloys at the core conditions.

Diffusion coefficients calculated using $4 \times 4 \times 2$ supercells and 10 ps simulation time are shown with open symbols. Blue squares: $\text{FeH}_{0.25}$ at ~ 260 GPa; red squares: $\text{FeH}_{0.25}$ at ~ 360 GPa; cyan triangles: $\text{FeO}_{0.0625}$ at ~ 260 GPa; orange triangles: $\text{FeO}_{0.0625}$ at ~ 360 GPa; green circles: $\text{FeC}_{0.0625}$ at ~ 260 GPa; pink circles: $\text{FeC}_{0.0625}$ at ~ 360 GPa. The convergence test results using $4 \times 4 \times 6$ supercell and 100 ps simulation time are labeled by crosses and bars. The results of convergence test are presented with yellow, magenta, and cyan symbols for $\text{FeH}_{0.25}$, $\text{FeO}_{0.0625}$, and $\text{FeC}_{0.0625}$, respectively.

[Source data](#)

Extended Data Fig. 4 Electronic conductivities of Fe and Fe alloys at 360 GPa with increasing temperature.

The electronic conductivities of Fe, $\text{FeH}_{0.25}$, $\text{FeO}_{0.0625}$, and $\text{FeC}_{0.0625}$ calculated by DFT + DMFT method are shown by black, blue, red, and

magenta symbols.

[Source data](#)

Extended Data Fig. 5 Calculated Poisson's ratios of FeH_{0.25}, FeC_{0.0625} and FeO_{0.0625} at various temperatures and 360 GPa.

Increasing temperature leads to obvious increases in Poisson's ratios approaching the value of the inner core (~0.44).

[Source data](#)

Extended Data Fig. 6

Calculated chemical potentials of X (X = H, O, C, S, and Si) in hcp-Fe with different configurations at 360 GPa and 0 K. The chemical potentials are shown with blue bars. The superscripts s and i denote the substitutional and interstitial defects. Separated(sub. + inter.) and correlated (dimer) configurations are noted.

[Source data](#)

Extended Data Fig. 7 The stability of interstitial and substitutional H, C and O in hcp-Fe under inner core conditions.

a, The MSD of Fe in Fe₆₀H₄ at 360 GPa and 5000 K; The relative formation energy of **b**, C and **c**, O at interstitial and substitutional site at 340 GPa and 360 GPa, respectively.

[Source data](#)

Extended Data Fig. 8 MSDs of Si, S, and Fe in FeSi_{0.0625} and FeS_{0.0625} at ~330 GPa and 3000 K.

MSDs of **a**, Si and Fe in $\text{FeSi}_{0.0625}$; **b**, S and Fe in $\text{FeS}_{0.0625}$. The MSDs of Si, S and Fe increase obviously with simulation time indicating a liquid state.

Source data

Extended Data Fig. 9 The structures of two-phase systems of Fe-H, Fe-O and Fe-C after the AIMD simulations.

These structures suggest the coexistence of solid and liquid Fe alloys. Pink, red, black and brown spheres represent H, O, C and Fe atoms, respectively.

Extended Data Table 1 Densities (ρ), elastic constants (C_{ij}), sound velocities (V_Φ , V_P and V_S), moduli (B and G) and Poisson's ratio of $\text{FeH}_{0.25}$, $\text{FeC}_{0.0625}$ and $\text{FeO}_{0.0625}$ at various temperatures and 360 GPa

Supplementary information

Supplementary Information

Supplementary Discussion, Figs. 1–4, Table 1 and References.

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**](#)

[**Source Data Extended Data Fig. 8**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

He, Y., Sun, S., Kim, D.Y. *et al.* Superionic iron alloys and their seismic velocities in Earth's inner core. *Nature* **602**, 258–262 (2022).
<https://doi.org/10.1038/s41586-021-04361-x>

- Received: 13 April 2020
- Accepted: 21 October 2021
- Published: 09 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04361-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/s41586-021-04361-x>

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

- Article
- [Published: 22 December 2021](#)

A species-level timeline of mammal evolution integrating phylogenomic data

- [Sandra Álvarez-Carretero](#)^{1,2 na1},
- [Asif U. Tamuri](#)^{3,4 na1},
- [Matteo Battini](#) [ORCID: orcid.org/0000-0001-6353-0138](#)⁵,
- [Fabrícia F. Nascimento](#)⁶,
- [Emily Carlisle](#)⁵,
- [Robert J. Asher](#)⁷,
- [Ziheng Yang](#) [ORCID: orcid.org/0000-0003-3351-7981](#)²,
- [Philip C. J. Donoghue](#) [ORCID: orcid.org/0000-0003-3116-7463](#)⁵ &
- [Mario dos Reis](#) [ORCID: orcid.org/0000-0001-9514-3976](#)¹

[Nature](#) volume **602**, pages 263–267 (2022)

- 11k Accesses
- 308 Altmetric
- [Metrics details](#)

Subjects

- [Molecular evolution](#)
- [Palaeontology](#)
- [Phylogenetics](#)
- [Statistical methods](#)

Abstract

High-throughput sequencing projects generate genome-scale sequence data for species-level phylogenies^{1,2,3}. However, state-of-the-art Bayesian methods for inferring timetrees are computationally limited to small datasets and cannot exploit the growing number of available genomes⁴. In the case of mammals, molecular-clock analyses of limited datasets have produced conflicting estimates of clade ages with large uncertainties^{5,6}, and thus the timescale of placental mammal evolution remains contentious^{7,8,9,10}. Here we develop a Bayesian molecular-clock dating approach to estimate a timetree of 4,705 mammal species integrating information from 72 mammal genomes. We show that increasingly larger phylogenomic datasets produce diversification time estimates with progressively smaller uncertainties, facilitating precise tests of macroevolutionary hypotheses. For example, we confidently reject an explosive model of placental mammal origination in the Palaeogene⁸ and show that crown Placentalia originated in the Late Cretaceous with unambiguous ordinal diversification in the Palaeocene/Eocene. Our Bayesian methodology facilitates analysis of complete genomes and thousands of species within an integrated framework, making it possible to address hitherto intractable research questions on species diversifications. This approach can be used to address other contentious cases of animal and plant diversifications that require analysis of species-level phylogenomic datasets.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

Subscribe

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Fig. 1: Summary of the Bayesian sequential subtree dating approach.



Fig. 2: Bayesian estimation of mammal divergence times.

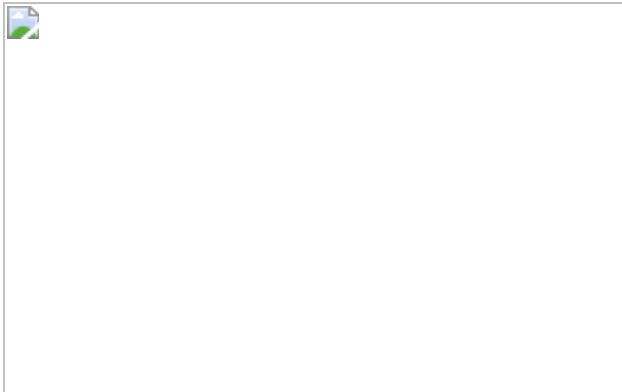


Fig. 3: Timetree of 4,705 mammal species.



Data availability

All data required to reproduce the analyses are available at
<https://doi.org/10.6084/m9.figshare.14885691>.

Code availability

A repository containing instructions to reproduce the analyses is available at http://github.com/sabifo4/mammals_dating and <https://doi.org/10.5281/zenodo.5736629>. The MCMCTree software and mcmc3r R package are freely available from <http://abacus.gene.ucl.ac.uk/software/paml.html> and <https://github.com/dosreislabs>, respectively.

References

1. Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
2. Feng, S. et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
3. Harvey, M. G. et al. The evolution of a tropical biodiversity hotspot. *Science* **370**, 1343–1348 (2020).
4. dos Reis, M., Donoghue, P. C. J. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80 (2016).
5. Meredith, R. W. et al. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
6. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494 (2019).
7. dos Reis, M. et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**, 3491–3500 (2012).
8. O’Leary, M. A. et al. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667 (2013).
9. dos Reis, M., Donoghue, P. C. & Yang, Z. Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biol. Lett.* **10**, 20131003 (2014).
10. Phillips, M. J. Geomolecular dating and the origin of placental mammals. *Syst. Biol.* **65**, 546–557 (2016).

11. Lewin, H. A. et al. Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl Acad. Sci. USA* **115**, 4325–4333 (2018).
12. Siepel, A. Challenges in funding and developing genomic software: roots and remedies. *Genome Biol.* **20**, 147 (2019).
13. Faria, N. R. et al. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014).
14. Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* **4**, 1232–1238 (2020).
15. Whelan, N. V. et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* **1**, 1737–1746 (2017).
16. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
17. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
18. Tao, Q., Tamura, K. & Kumar, S. in *The Molecular Evolutionary Clock: Theory and Practice* (ed. Ho, S. Y. W.) 197–219 (Springer, 2020).
19. Thorne, J. L. & Kishino, H. in *Statistical Methods in Molecular Evolution* (ed. Nielsen, R.) 235–256 (Springer, 2005).
20. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
21. Inoue, J., Donoghue, P. C. J. & Yang, Z. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89 (2010).

22. dos Reis, M., Zhu, T. & Yang, Z. The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *Syst. Biol.* **63**, 555–565 (2014).
23. Rannala, B. & Yang, Z. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466 (2007).
24. Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl Acad. Sci. USA* **100**, 1056–1061 (2003).
25. Hasegawa, M., Thorne, J. L. & Kishino, H. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* **78**, 267–283 (2003).
26. Alroy, J. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* **48**, 107–118 (1999).
27. Benton, M. J. Early origins of modern birds and mammals: Molecules vs. morphology. *Bioessays* **21**, 1043–1051 (1999).
28. Hunter, J. P. & Janis, C. M. Spiny Norman in the Garden of Eden? Dispersal and early biogeography of Placentalia. *J. Mamm. Evol.* **13**, 89–123 (2006).
29. Luo, Z. X. Transformation and diversification in early mammal evolution. *Nature* **450**, 1011–1019 (2007).
30. Cooper, A. & Fortey, R. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* **13**, 151–156 (1998).
31. Archibald, J. D. & Deutschman, D. H. Quantitative analysis of the timing of the origin and diversification of extant placental orders. *J. Mamm. Evol.* **8**, 107–124 (2001).
32. Murphy, W. J., Foley, N. M., Bredemeyer, K. R., Gatesy, J. & Springer, M. S. Phylogenomics and the genetic architecture of the placental mammal radiation. *Annu. Rev. Anim. Biosci.* **9**, 29–53 (2021).

33. Tarver, J. E. et al. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* **8**, 330–344 (2016).
34. Liu, L. et al. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proc. Natl Acad. Sci. USA* **114**, E7282–E7290 (2017).
35. dos Reis, M. et al. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst. Biol.* **67**, 594–615 (2018).
36. dos Reis, M. & Yang, Z. The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* **51**, 30–43 (2013).
37. dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
38. Battistuzzi, F. U., Billing-Ross, P., Paliwal, A. & Kumar, S. Fast and slow implementations of relaxed-clock methods show similar patterns of accuracy in estimating divergence times. *Mol. Biol. Evol.* **28**, 2439–2442 (2011).
39. Donoghue, P. C. J. & Yang, Z. The evolution of methods for establishing evolutionary timescales. *Philos. Trans. R. Soc. Lond. B* **371**, 20160020 (2016).
40. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160 (2011).
41. Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998).
42. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710

(2006).

43. Politis, D. N. & Romano, J. P. The stationary bootstrap. *J. Am. Stat. Assoc.* **89**, 1303–1313 (1994).
44. Nishihara, H., Maruyama, S. & Okada, N. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc. Natl Acad. Sci. USA* **106**, 5235–5240 (2009).
45. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
46. Bininda-Emonds, O. R. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
47. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
48. Zwart, S. P. The ecological impact of high-performance computing in astrophysics. *Nat. Astron.* **4**, 819–822 (2020).
49. Springer, M. S., Stanhope, M. J., Madsen, O. & de Jong, W. W. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* **19**, 430–438 (2004).
50. Kinsella, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011**, bar030 (2011).
51. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
53. Springer, M. S. & Gatesy, J. On the importance of homology in the age of phylogenomics. *Syst. Biodivers.* **16**, 210–228 (2018).

54. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
55. dos Reis, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
56. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
57. Löytynoja, A., Vilella, A. J. & Goldman, N. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics* **28**, 1684–1691 (2012).
58. Mitchell, K. J. et al. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Mol. Biol. Evol.* **31**, 2322–2330 (2014).
59. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
60. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
61. Hasegawa, M., Yano, T.-A. & Kishino, H. A new molecular clock of mitochondrial DNA and the evolution of hominoids. *Proc. Japan Acad. B* **60**, 95–98 (1984).
62. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
63. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
64. dos Reis, M. & Yang, Z. in *Evolutionary Genomics: Statistical and Computational Methods* (ed. Anisimova, M.) 309–330 (Springer,

2019).

65. Rannala, B., Zhu, T. & Yang, Z. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. *Mol Biol Evol.* **29**, 325–335 (2012).
66. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1–106 (2015).
67. Parham, J. F. et al. Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359 (2012).
68. Warnock, R. C. M., Parham, J. F., Joyce, W. G., Lyson, T. R. & Donoghue, P. C. J. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B* **282**, 20141013 (2015).
69. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
70. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
71. RStan: the R interface to Stan. R package. <https://mc-stan.org> (Stan Development Team, 2020).
72. Puttick, M. N. MCMCTreeR: functions to prepare MCMCTree analyses and visualize posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
73. Azzalini, A. *The R package ‘sn’: the skew-normal and related distributions such as the skew-t.* <http://azzalini.stat.unipd.it/SN/> (2019).

Acknowledgements

We thank J. Gilbert and C. G. Faulkes for help with the Rodentia subtrees. This work used computing resources from Queen Mary's Apocrita HPC and

University College London Myriad HPC facilities. This work was supported by Biotechnology and Biological Sciences Research Council, UK, awards BB/T01282X/1, BB/T012951/1 and BB/T012773/1.

Author information

Author notes

1. These authors contributed equally: Sandra Álvarez-Carretero, Asif U. Tamuri

Affiliations

1. School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK

Sandra Álvarez-Carretero & Mario dos Reis

2. Department of Genetics, Evolution and Environment, University College London, London, UK

Sandra Álvarez-Carretero & Ziheng Yang

3. Centre for Advanced Research Computing, University College London, London, UK

Asif U. Tamuri

4. EMBL-EBI, Wellcome Genome Campus, Hinxton, UK

Asif U. Tamuri

5. School of Earth Sciences, University of Bristol, Bristol, UK

Matteo Battini, Emily Carlisle & Philip C. J. Donoghue

6. MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK

Fabrícia F. Nascimento

7. Department of Zoology, University of Cambridge, Cambridge, UK

Robert J. Asher

Contributions

M.d.R. conceived the work. M.d.R., Z.Y., P.C.J.D., S.Á.-C. and A.U.T. designed the analysis. S.Á.-C., A.U.T., R.J.A., P.C.J.D., M.B., E.C. and F.F.N. compiled, processed and verified the molecular and fossil data. S.Á.-C., A.U.T. and M.d.R. analysed the data. M.d.R. and P.C.J.D. wrote the paper with input from all authors.

Corresponding authors

Correspondence to [Philip C. J. Donoghue](#) or [Mario dos Reis](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Olaf Bininda-Emonds, Xing-Xing Shen and the other, anonymous reviewers for their contribution to the peer review of this work.

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 Comparison of prior and posterior times.

a, Prior distribution of node ages generated by MCMC sampling without the molecular alignment. **b**, Posterior distribution of node ages when the 72-genome alignment is included during MCMC sampling. In both **a** and **b**, nodes are plotted at their posterior mean ages. The blue horizontal bars indicate the 95% credibility intervals of node ages.

Extended Data Fig. 2 Impact of fossil calibration strategies on node age estimates.

The posterior of node ages for the 72-taxon phylogeny is estimated using two additional fossil calibration strategies (y-axis) and plotted against the main estimates using best practice in calibration choice³⁹ (x-axis). In all cases the fossil minima are the same, but the calibration maxima changes. In the first strategy (black dots), calibration densities are narrow and close to the fossil ages. A truncated-Cauchy with a short tail (using $p = 0$ and $c = 0.001$, which extends the tail to about 110% of the fossil age) is used²¹. This strategy assumes the fossil record is a good indicator of the true node ages. In the second strategy (red dots), a truncated-Cauchy with a heavy tail (using $p = 0.1$ and $c = 1$, which extends the tail to over 900% of the fossil age) is used²¹. This strategy ignores the presence and absence of stem and sister groups, their palaeoecology, palaeobiogeography, and comparative taphonomy³⁹; and instead, assumes the node ages can be arbitrarily old. Dots are plotted at the posterior mean ages and vertical and horizontal bars indicate 95% CIs. The solid line is the $x = y$ line. The dashed lines are the regression lines for the corresponding data points.

Supplementary information

Supplementary Information

This file contains Supplementary Information, including Supplementary Figs. 1–11, Tables 1–13, annex and additional references.

Reporting Summary

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Álvarez-Carretero, S., Tamuri, A.U., Battini, M. *et al.* A species-level timeline of mammal evolution integrating phylogenomic data. *Nature* **602**, 263–267 (2022). <https://doi.org/10.1038/s41586-021-04341-1>

- Received: 06 July 2021
- Accepted: 13 December 2021
- Published: 22 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04341-1>

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/s41586-021-04341-1>

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

- Article
- [Published: 02 February 2022](#)

Autism genes converge on asynchronous development of shared neuron classes

- [Bruna Paulsen](#) [ORCID: orcid.org/0000-0001-5304-0610](#)^{1,2} [na1](#),
- [Silvia Velasco](#) [ORCID: orcid.org/0000-0003-0638-0932](#)^{1,2,16} [na1](#),
- [Amanda J. Kedaigle](#) [ORCID: orcid.org/0000-0001-6156-5046](#)^{1,2,3} [na1](#),
- [Martina Pigoni](#) [ORCID: orcid.org/0000-0002-7896-6978](#)^{1,2} [na1](#),
- [Giorgia Quadrato](#) [ORCID: orcid.org/0000-0003-0717-4874](#)^{1,17} [nAff11](#),
- [Anthony J. Deo](#)^{1,2} [nAff18](#) [nAff19](#) [nAff20](#),
- [Xian Adiconis](#)^{2,3},
- [Ana Uzquiano](#)^{1,2},
- [Rafaela Sartore](#)^{1,2},
- [Sung Min Yang](#)^{1,2},
- [Sean K. Simmons](#)^{2,3},
- [Panagiotis Symvoulidis](#) [ORCID: orcid.org/0000-0002-1396-3426](#)⁴,
- [Kwanho Kim](#) [ORCID: orcid.org/0000-0002-4448-3193](#)^{1,2,3},
- [Kalliopi Tsafou](#)²,
- [Archana Podury](#)^{4,5},
- [Catherine Abbate](#)^{1,2},
- [Ashley Tucewicz](#)^{1,2},
- [Samantha N. Smith](#)^{1,2},
- [Alexandre Albanese](#)⁶,
- [Lindy Barrett](#) [ORCID: orcid.org/0000-0002-3229-4774](#)^{1,2},
- [Neville E. Sanjana](#)² [nAff21](#) [nAff22](#),
- [Xi Shi](#)²,
- [Kwanghun Chung](#) [ORCID: orcid.org/0000-0002-8167-3340](#)^{6,7,21},

- [Kasper Lage](#) ORCID: [orcid.org/0000-0001-6827-6239^{2,8}](https://orcid.org/0000-0001-6827-6239),
- [Edward S. Boyden](#)^{4,5,9,10,11,12,13,14},
- [Aviv Regev](#)^{3,15} nAff23,
- [Joshua Z. Levin](#) ORCID: [orcid.org/0000-0002-0170-3598^{2,3}](https://orcid.org/0000-0002-0170-3598) &
- [Paola Arlotta](#) ORCID: [orcid.org/0000-0003-2184-2277^{1,2}](https://orcid.org/0000-0003-2184-2277)

[Nature](#) volume **602**, pages 268–273 (2022)

- 9910 Accesses
- 197 Altmetric
- [Metrics details](#)

Subjects

- [Autism spectrum disorders](#)
- [Developmental neurogenesis](#)
- [Neuroscience](#)
- [Stem cells](#)

Abstract

Genetic risk for autism spectrum disorder (ASD) is associated with hundreds of genes spanning a wide range of biological functions^{1,2,3,4,5,6}. The alterations in the human brain resulting from mutations in these genes remain unclear. Furthermore, their phenotypic manifestation varies across individuals^{7,8}. Here we used organoid models of the human cerebral cortex to identify cell-type-specific developmental abnormalities that result from haploinsufficiency in three ASD risk genes—*SUV420H1* (also known as *KMT5B*), *ARID1B* and *CHD8*—in multiple cell lines from different donors, using single-cell RNA-sequencing (scRNA-seq) analysis of more than 745,000 cells and proteomic analysis of individual organoids, to identify phenotypic convergence. Each of the three mutations confers asynchronous development of two main cortical neuronal lineages—γ-aminobutyric-acid-releasing (GABAergic) neurons and deep-layer excitatory projection

neurons—but acts through largely distinct molecular pathways. Although these phenotypes are consistent across cell lines, their expressivity is influenced by the individual genomic context, in a manner that is dependent on both the risk gene and the developmental defect. Calcium imaging in intact organoids shows that these early-stage developmental changes are followed by abnormal circuit activity. This research uncovers cell-type-specific neurodevelopmental abnormalities that are shared across ASD risk genes and are finely modulated by human genomic context, finding convergence in the neurobiological basis of how different risk genes contribute to ASD pathology.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: The *SUV420H1^{+/−}* genotype induces the asynchronous generation of GABAergic neurons and deep-layer projection neurons, and changes in circuit activity.

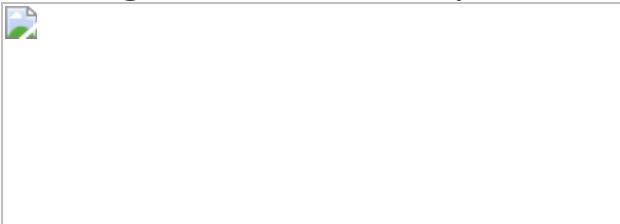


Fig. 2: The *ARID1B^{+/−}* genotype induces asynchronous generation of GABAergic neurons and deep-layer projection neurons.

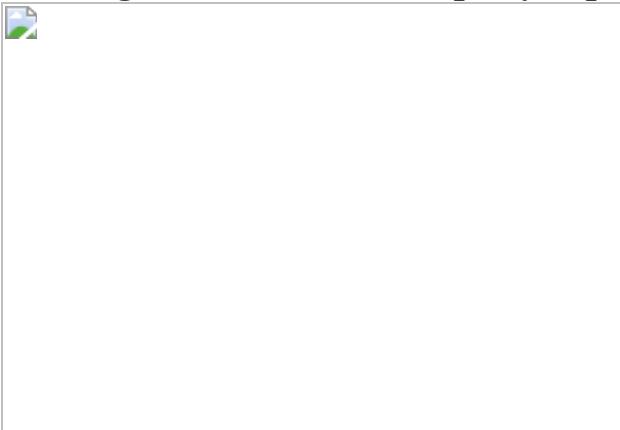


Fig. 3: The *CHD8^{+/−}* genotype leads to asynchronous generation of GABAergic interneurons.

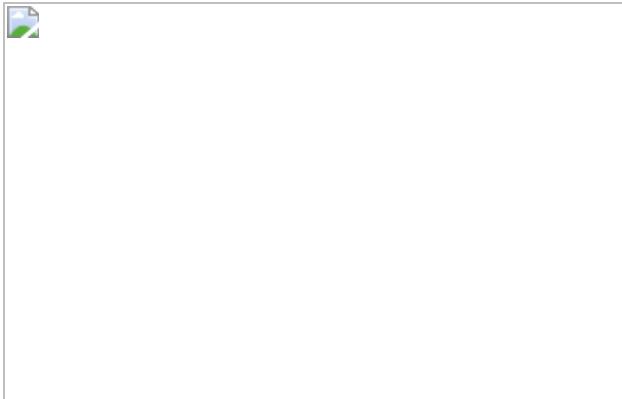
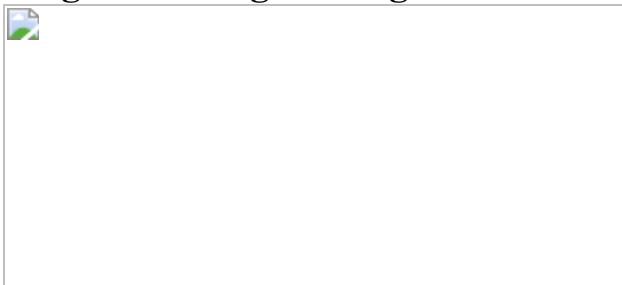


Fig. 4: The *SUV420H1^{+/−}*, *ARID1B^{+/−}* and *CHD8^{+/−}* genotypes act through distinct gene targets.



Data availability

Read-level data from scRNA-seq and scATAC-seq, along with proteomics data, supporting the findings of this study have been deposited in a controlled access repository at <https://www.synapse.org> with accession number project ID syn26346373 Count-level data and metadata have been deposited at the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1129). The electrophysiology materials and data are available from the corresponding authors on request. Public data used in this paper include the GRCh38 human reference genome and the EnsDb.Hsapiens.v86 annotation package.

Code availability

The code used for data analysis is available at GitHub (<https://github.com/AmandaKedaigle/mutated-brain-organoids>).

References

1. Lord, C. et al. Autism spectrum disorder. *Nat. Rev. Dis. Primers* **6**, 5 (2020).
2. Rosenberg, R. E. et al. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163**, 907–914 (2009).
3. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
4. Ruzzo, E. K. et al. Inherited and de novo genetic risk for autism impacts shared networks. *Cell* **178**, 850–866 (2019).
5. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
6. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584 (2020).
7. Cooper, D., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
8. Zlotogora, J. Penetrance and expressivity in the molecular age. *Genet. Med.* **5**, 347–352 (2003).
9. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
10. de Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
11. Stessman, H. A. F. et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).

12. Yuen, R. K. C. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
13. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
14. Bernier, R. et al. Disruptive *CHD8* mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
15. Faundes, V. et al. Histone lysine methylases and demethylases in the landscape of human developmental disorders. *Am. J. Hum. Genet.* **102**, 175–187 (2018).
16. Vals, M. et al. Coffin-Siris syndrome with obesity, macrocephaly, hepatomegaly and hyperinsulinism caused by a mutation in the *ARID1B* gene. *Eur. J. Hum. Genet.* **22**, 1327–1329 (2014).
17. Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
18. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
19. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
20. Wickramasekara, R. N. & Stessman, H. A. F. Histone 4 lysine 20 methylation: a case for neurodevelopmental disease. *Biology* **8**, 11 (2019).
21. Garaschuk, O., Linn, J., Eilers, J. & Konnerth, A. Large-scale oscillatory calcium waves in the immature cortex. *Nat. Neurosci.* **3**, 452–459 (2000).

22. Adelsberger, H., Garaschuk, O. & Konnerth, A. Cortical calcium waves in resting newborn mice. *Nat. Neurosci.* **8**, 988–990 (2005).
23. Wang, Z.-J. et al. Autism risk gene *KMT5B* deficiency in prefrontal cortex induces synaptic dysfunction and social deficits via alterations of DNA repair and gene transcription. *Neuropsychopharmacology* **46**, 1617–1626 (2021).
24. Villa, C. E. et al. CHD8 haploinsufficiency alters the developmental trajectories of human excitatory and inhibitory neurons linking autism phenotypes with transient cellular defects. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.11.26.399469> (2020).
25. Wang, P. et al. CRISPR/Cas9-mediated heterozygous knockout of the autism gene *CHD8* and characterization of its transcriptional networks in cerebral organoids derived from iPS cells. *Mol. Autism* **8**, 11 (2017).
26. Tuncbag, N. et al. Network-based interpretation of diverse high-throughput datasets through the Omics Integrator software package. *PLoS Comput. Biol.* **12**, e1004879 (2016).
27. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
28. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
29. Gogolla, N. et al. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J. Neurodev. Disord.* **1**, 172–181 (2009).
30. Dani, V. S. et al. Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA* **102**, 12560–12565 (2005).

31. Mariani, J. et al. FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell* **162**, 375–390 (2015).
32. Marchetto, M. C. et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* **22**, 820–835 (2017).
33. Adhya, D. et al. Atypical neurogenesis in induced pluripotent stem cells from autistic individuals. *Biol. Psychiatry* **89**, 486–496 (2020).
34. Wade, A. A., Lim, K., Catta-Preta, R. & Nord, A. S. Common CHD8 genomic targets contrast with model-specific transcriptional impacts of CHD8 haploinsufficiency. *Front. Mol. Neurosci.* **11**, 481 (2019).
35. Moffat, J. J., Smith, A. L., Jung, E. M., Ka, M. & Kim, W. Y. Neurobiology of ARID1B haploinsufficiency related to neurodevelopmental and psychiatric disorders. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-021-01060-x> (2021).
36. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
37. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
38. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).
39. Chen, A. E. et al. Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).
40. Church, G. M. The personal genome project. *Mol. Syst. Biol.* **1**, 2005.0030 (2005).

41. Velasco, S., Paulsen, B. & Arlotta, P. Highly reproducible human brain organoids recapitulate cerebral cortex cellular diversity. *Protoc. Exchange* <https://doi.org/10.21203/rs.2.9542/v1> (2019).
42. Lovell-Badge, R. et al. ISSCR guidelines for stem cell research and clinical translation: the 2021 update. *Stem Cell Rep.* **16**, 1398–1408 (2021).
43. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
44. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
45. Mangeot, P. E. et al. Genome editing in primary cells and in vivo using viral-derived nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
46. Park, Y.-G. et al. Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat. Biotechnol.* **37**, 73–83 (2019).
47. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
48. Ohgane, K. Quantification of gel bands by an Image J macro, band/peak quantification tool. *protocols.io* <https://doi.org/10.17504/protocols.io.7vghn3w> (2019).
49. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRxiv* <https://doi.org/10.1101/061507> (2017).
50. Müller, J. et al. High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels. *Lab Chip* **15**, 2767–2780 (2015).

51. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wisniewski, J. R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
52. Wiśniewski, J. R. Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols. *Anal. Chem.* **88**, 5438–5443 (2016).
53. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**, 49–54 (1999).
54. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2018).
55. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7**, 40–44 (2008).
56. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
57. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
59. Tuncbag, N. et al. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *J. Comput. Biol.* **20**, 124–136 (2013).
60. Akhmedov, M. et al. PCSF: an R-package for network-based interpretation of high-throughput data. *PLoS Comput. Biol.* **13**,

e1005694 (2017).

61. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
62. Yoon, S. et al. GScluster: network-weighted gene-set clustering analysis. *BMC Genom.* **20**, 352 (2019).
63. Quadrato, G., Sherwood, J. L. & Arlotta, P. Long term culture and electrophysiological characterization of human brain organoids. *Protoc. Exchange* <https://doi.org/10.1038/protex.2017.049> (2017).
64. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
65. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
66. McCarthy, D. J. et al. Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat. Methods* **17**, 414–421 (2020).
67. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
68. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
69. Pardy, C. mpmi: mixed-pair mutual information estimators (2020).
70. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
71. Fonseka, C. Y. et al. Mixed-effects association of single cells identifies an expanded effector CD4⁺ T cell subset in rheumatoid arthritis. *Sci. Transl. Med.* **10**, eaaq0305 (2018).

72. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
73. 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).
74. Lun, A. T. L. & Marioni, J. C. Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* **18**, 451–464 (2017).
75. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
76. *Nuclei Isolation from Mouse Brain Tissue for Single Cell ATAC Sequencing Rev B* (10x Genomics, 2019).
77. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
78. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
79. Rainer, J. EnsDb.Hsapiens.v86: ensembl based annotation package (2017).
80. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

Acknowledgements

We thank J. R. Brown (from the P.A. laboratory) for input and assistance in editing the manuscript and all of the members of the Arlotta laboratory for discussions; M. Daly, E. Robinson and B. Neale for discussions on autism

genetics; T. Nguyen (from the G.Q. laboratory) for support with organoid generation; X. Jin (from the P.A. laboratory) for helping with designing and sequencing edited lines; A. Shetty (from the P.A. laboratory) for help with scRNA-seq cell type classification; N. Haywood for help with scRNA-seq experiments; D. Di Bella (from the P.A. laboratory) for help with final edits of the manuscript; S. Andreadis and S. Getz (from the P.A. laboratory) for help with editing images; F. Zhang and J. Pan for supporting the creation of the HUES66 *CHD8*-mutant line; members of the Talkowski laboratory for the GM08330 line; members of the Cohen laboratory for the Mito210 line; members of the Ricci laboratory for providing nanoblades for the generation of *SUV420H1* edited cell lines; L. M. Daheron at the Harvard Stem Cell Facility for expanding edited lines; and B. Budnik at the Harvard Center for Mass Spectrometry for assisting with proteomics experiments. This work was supported by grants from the Stanley Center for Psychiatric Research, the Broad Institute of MIT and Harvard, the National Institutes of Health (R01-MH112940 to P.A. and J.Z.L.; P50MH094271, U01MH115727 and 1RF1MH123977 to P.A.), the Klarman Cell Observatory to J.Z.L. and A.R., and the Howard Hughes Medical Institute to A.R. A.R. was a Howard Hughes Medical Institute and a Koch Institute extramural member while conducting this study. The HUES66 *CHD8*-mutant line was created with support from the Simons Foundation (346073 to F. Zhang) and the National Institutes of Health (MH099448 to J. Pan).

Author information

Author notes

1. Giorgia Quadrato

Present address: Howard Hughes Medical Institute, MIT, Cambridge, MA, USA

2. Anthony J. Deo

Present address: Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

3. Anthony J. Deo

Present address: Department of Psychiatry, Rutgers-Robert Wood Johnson Medical School, Piscataway, NJ, USA

4. Anthony J. Deo

Present address: Rutgers University Behavioral Health Care, Piscataway, NJ, USA

5. Neville E. Sanjana

Present address: New York Genome Center, New York, NY, USA

6. Neville E. Sanjana

Present address: Department of Biology, New York University, New York, NY, USA

7. Aviv Regev

Present address: Genentech, South San Francisco, CA, USA

8. These authors contributed equally: Bruna Paulsen, Silvia Velasco, Amanda J. Kedaigle, Martina Pignoni

Affiliations

1. Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

Bruna Paulsen, Silvia Velasco, Amanda J. Kedaigle, Martina Pignoni, Giorgia Quadrato, Anthony J. Deo, Ana Uzquiano, Rafaela Sartore, Sung Min Yang, Kwanho Kim, Catherine Abbate, Ashley Tucewicz, Samantha N. Smith, Lindy Barrett & Paola Arlotta

2. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Bruna Paulsen, Silvia Velasco, Amanda J. Kedaigle, Martina Pignoni, Anthony J. Deo, Xian Adiconis, Ana Uzquiano, Rafaela Sartore, Sung Min Yang, Sean K. Simmons, Kwanho Kim, Kalliopi Tsafou, Catherine Abbate, Ashley Tucewicz, Samantha N. Smith, Lindy Barrett, Neville E. Sanjana, Xi Shi, Kasper Lage, Joshua Z. Levin & Paola Arlotta

3. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Amanda J. Kedaigle, Xian Adiconis, Sean K. Simmons, Kwanho Kim, Aviv Regev & Joshua Z. Levin

4. MIT Center for Neurobiological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Panagiotis Symvoulidis, Archana Podury & Edward S. Boyden

5. Harvard-MIT Health Sciences & Technology Program (HST), Harvard Medical School, Boston, MA, USA

Archana Podury & Edward S. Boyden

6. Institute for Medical Engineering and Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Alexandre Albanese & Kwanghun Chung

7. Picower Institute for Learning and Memory, Departments of Chemical Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Kwanghun Chung

8. Department of Surgery and Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

Kasper Lage

9. McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Edward S. Boyden

10. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Edward S. Boyden

11. Howard Hughes Medical Institute, MIT, Cambridge, MA, USA

Edward S. Boyden

12. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Edward S. Boyden

13. Department of Media Arts and Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Edward S. Boyden

14. Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Edward S. Boyden

15. Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Aviv Regev

16. Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia

Silvia Velasco

17. Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Giorgia Quadrato

18. New York Genome Center, New York, NY, USA

Kwanghun Chung

Contributions

P.A., B.P., S.V., A.J.K., M.P., and G.Q. conceived the experiments. A.J.D. designed *SUV420H1* and *ARID1B* gRNAs and generated the Mito210 and Mito294 *SUV420H1*-edited lines with B.P.; L.B. generated the *ARID1B*-edited lines. N.E.S. and X.S. designed *CHD8* gRNAs and generated the HUES66 *CHD8* line. A.J.D. and B.P. generated the Mito210 and Mito294 *CHD8*-edited lines. S.V., B.P., M.P., R.S., C.A., A.T and S.N.S. generated, cultured and characterized all of the organoids used in this study and P.A. supervised their work. X.A. performed scRNA-seq experiments with help from B.P., S.V., M.P., R.S. and G.Q. under the supervision of P.A. and J.Z.L.; A.J.K., K.K., S.K.S. and J.Z.L. performed scRNA-seq analysis and J.Z.L. and A.R. supervised their work. S.V., B.P., M.P., R.S., A.U., G.Q. and A.J.K. worked on cell type assignments and data analysis. K.T., M.P. and A.J.K. performed proteomic analysis, supervised by K.L.; S.M.Y., P.S. and A.P. performed the calcium imaging experiments and analysis, supervised by E.S.B. and P.A.; S.M.Y. and R.S. performed the MEA recordings and analysis supervised by P.A.. A.A. performed whole-organoid imaging under the supervision of K.C. P.A., B.P., S.V., A.J.K. and M.P. wrote the manuscript with contributions from all of the authors. All of the authors read and approved the final manuscript.

Corresponding authors

Correspondence to [Silvia Velasco](#) or [Paola Arlotta](#).

Ethics declarations

Competing interests

P.A. is a SAB member at Herophilus, Foresite Labs, and Rumi Scientific, a consultant for the New York Stem cell Foundation, and is a co-founder of esalius Therapeutics. A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and, until 31 August 2020, was a SAB member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and Thermo Fisher Scientific. From 1 August 2020, A.R. is an employee of Genentech. N.E.S. is an advisor to Vertex and Qiagen.

Peer review information

Nature thanks Prisca Liberali, Annie Vogel-Ciernia and Irina Voineagu for their contribution to the peer review of this work.

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 Cortical organoids cultured for one, three and six months generate the cellular diversity of the human cerebral cortex with high organoid-to-organoid reproducibility.](#)

a, scRNA-seq and immunohistochemistry analysis of organoids cultured for one month (32 d.i.v.), three months (98 d.i.v.), and six months (190 d.i.v.). Left, *t*-SNE plots ($n = 3$ organoids per timepoint, co-clustered). Cells are coloured by cell-type. Right, immunohistochemistry for specific markers. Neural progenitor marker SOX2 (magenta) and postmitotic neuronal marker TBR1 (green) are shown at one month. CPN marker SATB2

(magenta) and CFuPN marker CTIP2 (green) are shown at three months. The astroglia markers S100B (magenta) and GFAP (green) are shown at six months. Below, schematic images of brain organoids in each timepoint. Scale bars are 100 µm. **b**, Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2), CFuPN (CTIP2), and CPN (SATB2) markers in GM08330 organoids at one, three, and six months. Scale bars: whole organoids (leftmost column), 200 µm; others, 50 µm. **c**, Immunohistochemistry for cell-type specific markers in Mito210 organoids, as in **b**. **d**, Top, *t*-SNE plots of the scRNA-seq data from individual replicates from three organoids at one month, three organoids at three months, and three organoids at six months from the GM08330 cell line shown in **b**. Bottom, bar charts showing the cell-type composition of each individual organoid. On top of the bar charts, mutual information (MI) scores between cell-type proportions and organoid identities are displayed. A MI score of 0 would indicate identical cell type proportions between organoids, while a score of 1 would indicate completely divergent profiles. In previous work, MI scores for endogenous brain datasets were reported to range from 0.008 to 0.064⁹. **e**, scRNA-seq data of organoids from the Mito210 cell line at one month (35 d.i.v.), three months (92 d.i.v.), and six months (178 d.i.v.), as in **d**. Organoids for the one and three month timepoints are the same as the control organoids in Extended Data Fig. [4f](#) and Extended Data Fig. [5b](#). **f**, Expression of selected marker genes used in cell-type identification. Violin plots show distribution of normalized expression in cells from GM08330 organoids at one, three and six months ($n = 3$ individual organoids per timepoint). **g**, Expression of marker genes in Mito210 organoids, as in **f**. Number of organoids used for each analysis can be found in the Methods under “Statistics and reproducibility”. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.

[Extended Data Fig. 2 Expression of selected ASD risk genes in cortical organoids cultured for one, three, and six months.](#)

a, *t*-SNE plots of 58,568 cells from nine organoids from the GM08330 cell line, shown in Extended Data Fig. [1d](#), after Harmony batch correction. Cells

are coloured according to cell-type (left) and timepoint (right). **b**, Gene set expression scores for a set of 102 genes associated with ASD risk⁶ across cell-types, in cells from **a**. Scores above 0 indicate enriched expression over similar sets of randomly chosen genes. **c**, t-SNE plots showing normalized expression of selected ASD risk genes in cells from **a**. **d**, Average expression of 102 genes associated with ASD risk across cell-types and timepoints in the GM08330 cell line. **e**, t-SNE plots of nine organoids from the Mito210 cell line, shown in Extended Data Fig. [1e](#), after Harmony batch correction. Cells are coloured according to cell-type (left) or timepoint (right). **f**, Gene set scores for the set of ASD risk genes as in **b**, in cells from **e**. Scores above 0 indicate higher expression than similar modules of randomly chosen genes. **g**, t-SNE plots showing normalized expression of selected ASD risk genes in cells from **e**. **h**, Expression of 102 genes associated with ASD risk across cell-types and timepoints in Mito210 cell line. RG, radial glia (aRG, oRG, and oRG/Astroglia), IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; EN, Excitatory neurons (CPN, CFuPN and PN); GABA IN, GABAergic interneurons.

Extended Data Fig. 3 Generation and characterization of SUV420H1, ARID1B, and CHD8 mutant organoids.

a, Protein domain structure of SUV420H1. Arrow indicates the region (N-domain) mutated in the Mito210, PGP1 and Mito294 parental lines (bottom). **b**, Protein domain structure of ARID1B. Arrow indicates the region before the ARID domain mutated in the Mito210 and Mito294 parental lines (bottom). **c**, Protein domain structure of CHD8. Arrows indicates the helicase C-terminal (HELC) domain mutated in the HUES66, H1, GM08330, Mito294 and Mito210 lines (bottom). **d–f**, Western blot analysis showing presence of SUV420H1 (**d**), ARID1B (**e**) and CHD8 (**f**) protein expression in control lines, and its reduction in the mutant lines. Molecular weight in kDa is shown on the left of the gel. H4K20me3, a hallmark of SUV420H1 activity, and total levels of histone H4 were also detected in control and in *SUV420H1*^{+/−} lines (**d**). ARID1B was not detectable in Mito294 even with a longer exposure of the blotted membrane (**e**, right). Asterisks indicate the bands used for quantification. Bottom, protein levels in control and mutant lines were quantified and normalized

for housekeeping genes β-Actin or GAPDH. For gel source data, see Supplementary Fig. 1. **g**, Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2) and CFuPN (CTIP2) markers in organoids at 35 d.i.v. derived from the Mito210 *SUV420H1*^{+/−}, Mito210 *ARID1B*^{+/−} and HUES66 *CHD8*^{+/−} and isogenic control cell lines. Scale bar, 300 μm. **h–j**, Size quantification of control and *SUV420H1*^{+/−} (**h**), *ARID1B*^{+/−} (**i**) and *CHD8*^{+/−} (**j**) organoids across lines and at different timepoints. The ratio of organoid size compared to the average of control organoids in each batch is plotted. Differentiation batch (b.) is indicated by the shade of the points. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and middle hinge is the median (50th). Both whiskers extends from the hinge to the largest or smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). *P*-values from a two-sided t-test, after Bonferroni adjustment within each mutation. Number of organoids and differentiations used for the measurement are summarized in Supplementary Table 2 and in the Methods under “Statistics and reproducibility”.

Extended Data Fig. 4 Cell-type composition of *SUV420H1*^{+/−} and isogenic control organoids.

a, Immunohistochemistry of Mito210 *SUV420H1*^{+/−} and control organoids cultured for one month (35 d.i.v.). Optical section from the middle of whole-organoid dataset. Scale bars are 500 μm. SOX2, a marker of neuronal progenitors, is shown in red, and nuclei (Syt016) are shown in blue. **b**, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 and GABAergic marker DLX2 in Mito294 control and *SUV420H1*^{+/−} organoids at one month (35 d.i.v.). Scale bars: 200 μm. **c–e**, scRNA-seq data from one month (Mito294 35d.i.v. (**c**), PGP1 35 d.i.v. (**d**) and Mito210 28 d.i.v., batch I (**e**)) control and *SUV420H1*^{+/−} organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell type proportions between control and mutant, based on logistic mixed models (see [Methods](#)) are shown. **f**, scRNA-seq data from Mito210 35 d.i.v. (batch II) control and *SUV420H1*^{+/−} organoids. Left top shows combined t-SNE

plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell-type, and the total number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. Immature deep-layer projection neuron populations are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **c–e**. **g**, Enriched GO terms for genes upregulated and downregulated in *SUV420H1^{+/−}* vs. control across lines. Genes were calculated using cells from the partition of interest. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes (“GeneRatio”). Colour indicates enrichment adjusted *p*-value. Numbers in parentheses along the y axis indicate the number of DEGs in that dataset. As control, we calculated GO term enrichment for 100 random gene sets of the same size sampled from genes expressed in each dataset, and found no significant enrichment of these terms (see [Methods](#)). Number of organoids used for each analysis can be found in the Methods under “Statistics and reproducibility”. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA N, GABAergic neurons.

[Extended Data Fig. 5 Cell-type composition, full pseudotime trajectories, and gene modules in *SUV420H1^{+/−}* and isogenic control organoids.](#)

a–c, scRNA-seq data from three month Mito294 89 d.i.v. (**a**), Mito210 92 d.i.v. batch I (**b**), and 90 d.i.v. batch II (**c**) control and *SUV420H1^{+/−}* organoids. Left top shows combined *t*-SNE plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. Cell-types of interest are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell type proportions between control and mutant, based on logistic mixed models (see [Methods](#)) are shown. **d**, Pseudotime

trajectory from the full dataset of Mito210 $SUV420HI^{+/-}$ 35 d.i.v. (batch II) and control organoids, calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. 1d. e, Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from Mito210 $SUV420HI^{+/-}$ and control organoids at 35 d.i.v. (batch II). Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted p -values show differences between control and mutant based on linear mixed models (see [Methods](#)). aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons; GABA N, GABAergic neurons.

[Extended Data Fig. 6 scATAC-seq analysis in \$SUV420HI^{+/-}\$ and isogenic control organoids.](#)

a, UMAPs of scATAC-seq data in Mito210 $SUV420HI^{+/-}$ and control organoids at one month (31 d.i.v., upper left) and three months (93 d.i.v., upper right), and coembedded UMAPs with scRNA-seq in Mito210 $SUV420HI^{+/-}$ and control organoids at one month (28 d.i.v., middle bottom left) and three months (90 d.i.v., middle bottom right). Number of nuclei per plot is indicated. b, Enriched GO terms for the nearest genes to regions with increased and decreased accessibility in $SUV420HI^{+/-}$ compared to control organoids. c, Genome tracks showing read coverage for representative regions with increased accessibility between $SUV420HI^{+/-}$ and control organoids. Scales for the y axes (normalized counts) are displayed on the top right. d, Genome tracks showing read coverage for representative regions with increased accessibility between Mito210 $SUV420HI^{+/-}$ and control organoids, split by cell-type. Scales for the y axes (normalized counts) are displayed on the top right. e, Top 5 *de novo* motifs enriched in the regions with altered accessibility in Mito210 $SUV420HI^{+/-}$ compared

to control organoids at one month (31 d.i.v.) and three months (93 d.i.v.), as calculated with HOMER (see [Methods](#)). Regions that showed increased accessibility in mutant compared to control organoids are in green and purple, while those with decreased accessibility are in red and blue. Transcription factors with known binding sites resembling the discovered motifs are shown.

[Extended Data Fig. 7 Neuronal spontaneous activity in *SUV420HI*^{+/−} and isogenic control organoids.](#)

a, Left, Representative image of a PGP1 *SUV420HI* organoid infected with SomaGCaMP6f2. Right, $\Delta F/F$ signal at the peak of a network burst. Scale bar: 100 μm . **b**, Heat map of calcium signal from individual cells (rows), showing the effect of 2 μM TTX. **c**, Top, representative trace of spontaneous calcium signal (corresponding to cell #3 in Fig. [1g](#)). Bottom, high magnification traces of calcium transients, displaying the difference in amplitude between the isolated event and the network burst (top), and normalized traces (bottom) showing their kinetics and the multiple peaks of the burst signal. **d**, MEA recordings. Top, Spatial configuration of recording electrodes. Middle, example raw traces for the numbered electrodes shown at the top, and the effect upon 2 μM TTX application. Yellow columns indicate the network bursts. Right, individual (grey) and average (colour) spike waveforms for each electrode. High magnification of the trace #4 showing the individual spikes (asterisk) during a burst event. Bottom, average spike waveform (right) in a unit of electrodes (left), extracted at the time points determined by the spikes in electrode #4. **e, f**, Synchronous network activity in human brain organoids. Heat map of cross-correlation index (**e**) and cross-correlogram against a reference signal (cell #135) for a representative recording. As a control, the signal of 10 cells were circularly shifted by a random phase and the cross-correlation was then calculated. In **f**, the average value was plotted, and the synchronous activations as well as the periodic bursting can be seen (“All cells” in red). **g**, Effect of NBQX on neuronal activity. Representative traces for individual cells were normalized (3 traces for *SUV420HI*^{+/−} are superimposed) and post-NBQX residual/isolated calcium transients are indicated by asterisks. **h**, Effect of NBQX on calcium signal. Heat map of $\Delta F/F$ signal for 15 representative

cells in control (top) and *SUV420H1*^{+/−} (bottom) organoids. **i, j**, Inter-spike interval (ISI) analysis for the network bursting. Relative frequency (top) and cumulative frequency distribution (bottom) of ISI for control and *SUV420H1*^{+/−} organoids. In **j**, two-sided Kolmogorov-Smirnov test ($n = 5$ organoids per genotype). Number of organoids used for each analysis can be found in the Methods under “Statistics and reproducibility”.

Extended Data Fig. 8 Cell-type composition, full pseudotime trajectories, and gene modules of *ARID1B*^{+/−} and isogenic control organoids.

a, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in Mito210 control and *ARID1B*^{+/−} organoids at one month (35 d.i.v.). Scale bars: 200 μm. **b, c**, scRNA-seq data from Mito210 one month (35 d.i.v. batch I in **b**, batch II in **c**) control and *ARID1B*^{+/−} organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see [Methods](#)) are shown. **d**, Immunohistochemistry for TBR1 (magenta) and DLX2 (green) in Mito210 control and *ARID1B*^{+/−} organoids at three months (90 d.i.v.). Scale bars: 100 μm. **e**, scRNA-seq data from Mito210 three months (90 d.i.v.) control and *ARID1B*^{+/−} organoids. Left top shows combined t-SNE plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell-type, and the number of cells per plot is indicated. Left bottom, t-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Left, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as shown in **b, c**. Two out of three mutant organoids show an increase in GABAergic interneurons, vs. zero out of three controls (adjusted *p* = 0.19, logistic mixed models). **f**, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in Mito210 control and *ARID1B*^{+/−} organoids at three months (90 d.i.v.). Three out of four mutant organoids contain DLX2-positive cells, while no DLX2 expression is detected in the four controls. Scale bars: 500 μm. **g**, scRNA-seq data from Mito294 one month (35 d.i.v.)

ARID1B^{+/−} and control organoids. Left top shows combined *t*-SNE plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids.

GABAergic neurons, newborn deep-layer projection neurons and immature deep-layer projection neuron populations are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **b**, **c**, **e**. **h**, Pseudotime trajectories from the full dataset of Mito210 *ARID1B*^{+/−} 35 d.i.v. *batch II* and control organoids, calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. [2c](#). **i**, Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from Mito210 *ARID1B1*^{+/−} and control organoids at 35 d.i.v. Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted *p*-values show differences between control and mutant based on linear mixed models (see [Methods](#)). Number of organoids used for each analysis can be found in the Methods under “Statistics and reproducibility”. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA NP, GABAergic neuron progenitors; GABA N, GABAergic neurons; GABA INP; GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.

Extended Data Fig. 9 Cell-type composition, immunohistochemistry, and full pseudotime trajectories and gene modules of *CHD8*^{+/−} and isogenic control HUES66 organoids.

a, b, scRNA-seq data from HUES66 3.5-month (109 d.i.v. **(a)**, batch I and 107 d.i.v. **(b)**, batch II) *CHD8*^{+/−} and control organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant

organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see [Methods](#)) are shown. **c**, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in HUES66 control and *CHD8^{+/−}* organoids at 3.5 months (107 d.i.v.). Scale bars: 100 μ m. **d**, Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2) and CFuPN (CTIP2) markers in HUES66 *CHD8^{+/−}* and control organoids at 3.5 months (107 d.i.v., top), and six months (190 d.i.v., bottom). Scale bars: whole organoids, 500 μ m; others, 100 μ m. **e**, scRNA-seq data from HUES66 *CHD8^{+/−}* and control organoids at six months (190 d.i.v.). Top left shows combined *t*-SNE plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell-type, and the number of cells per plot is indicated. Top right, *t*-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Bottom, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **a**, **b**. **f**, Immunohistochemistry for the post mitotic neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in HUES66 control and *CHD8^{+/−}* organoids at six months (190 d.i.v.). Scale bars: 100 μ m. **g**, Pseudotime trajectories from the full dataset of HUES66 batch I *CHD8^{+/−}* and control organoids at 109 d.i.v., calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. [3c](#). **h**, Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from HUES66 *CHD8^{+/−}* and control organoids at 109 d.i.v. Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted *p*-values show differences between control and mutant based on linear mixed models (see [Methods](#)). Number of organoids used for each analysis can be found in the Methods under “Statistics and reproducibility”. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.

Extended Data Fig. 10 Bulk RNA-seq and scRNA-seq of *CHD8*^{+/−} and isogenic control organoids from multiple cell lines.

a, Bulk RNA-seq data from HUES66, GM83330 and H1 35 d.i.v. organoids. Enriched GO terms for genes differentially expressed in *CHD8*^{+/−} vs. control organoids. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes (“GeneRatio”). Colour indicates enrichment adjusted *p*-value. Numbers in parentheses along the y axis indicate the number of DEGs in that dataset. **b–d**, scRNA-seq data from control and *CHD8*^{+/−} organoids at 3.5 months (GM83330 108 d.i.v., batch I (**b**), GM83330 108 d.i.v., batch II (**c**) and H1 105 d.i.v. (**d**)). Left top shows combined *t*-SNE plots of control and mutant organoids ($n = 3$ single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see [Methods](#)) are shown. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons; GABA N, GABAergic neurons.

Extended Data Fig. 11 Convergent differential expressed genes for the three mutations.

a, Log fold change of all genes which showed significant change (adjusted $p < 0.05$) in all three of the 1 month datasets: Mito210 *SUV420H1*^{+/−} 35 d.i.v., Mito210 *ARID1B*^{+/−} 35 d.i.v., and HUES66 *CHD8*^{+/−} 35 d.i.v. DEGs were calculated using all cells as a pseudobulk for Mito210 *SUV420H1*^{+/−} and Mito210 *ARID1B*^{+/−}. **b**, Differential expression of all 102 genes

associated with ASD risk⁶ in the three datasets Mito210 *SUV420HI*^{+/−} 35 d.i.v., Mito210 *ARID1B*^{+/−} 35 d.i.v. and in HUES66 *CHD8*^{+/−} 35 d.i.v. compared to relative controls. Expression of risk genes was calculated using all cells (pseudobulk) for Mito210 *SUV420HI*^{+/−} and Mito210 *ARID1B*^{+/−}. Boxes are coloured according to $-\log_{10}(\text{adjusted } p \text{ value})$ according to whether they are upregulated (purple), or downregulated (turquoise) in mutant vs. control. Genes are ordered according to hierarchical clustering (using Euclidean distance) of those values. **c, d**, Enriched GO terms for genes upregulated (**c**) and downregulated (**d**) in mutant vs. control. Genes were calculated using the cells as in **a, b**. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes (“GeneRatio”). Colour indicates enrichment adjusted *p*-value. Numbers in parentheses along the x axis indicate the number of DEGs in that dataset.

Extended Data Fig. 12 Convergent differentially expressed proteins for the three mutations.

a–c, Volcano plot showing fold change versus adjusted *p*-value of measured proteins in MS experiments on Mito210 *SUV420HI*^{+/−} (**a**), Mito210 *ARID1B*^{+/−} (**b**), and HUES66 *CHD8*^{+/−} (**c**) vs. control organoids at 35 d.i.v. ($n = 4$ single organoids per genotype for *SUV420HI*, 4 controls and 5 mutants for *ARID1B*, and $n = 3$ single organoids per genotype for *CHD8*). To detect statistically significant differential protein abundance between conditions a moderated t-test was performed (see [Methods](#), FDR threshold of 0.1). Significant DEPs are shown in red (FDR < 0.1). **d–f**, Selected enriched GO terms for DEPs in Mito210 *SUV420HI*^{+/−} (**d**), Mito210 *ARID1B*^{+/−} (**e**), and HUES66 *CHD8*^{+/−} (**f**) vs. control organoids cultured for 35 d.i.v. GO terms and KEGG pathways were calculated using the GSEA software (see [Methods](#)) and FDR q-values < 0.05 were considered statistically significant. **g**, Protein-protein interaction network using the top 50 DEPs from the three sets of mutant *versus* control organoids, created using the prize-collecting Steiner forest algorithm (see [Methods](#)). Protein nodes are coloured by the mutant in which they were differentially expressed. Gray nodes indicate “Steiner nodes”, proteins that did not result from any screen but were included by the algorithm to connect DEPs. Lines

between nodes indicate physical protein-protein interactions from the STRING database, where line thickness correlates with interaction confidence. Subclusters of the network and significantly enriched terms for those subclusters are highlighted with gray rectangles and black text. **h**, Protein set distances between pairs of differentially expressed protein sets. For each pair of mutations, a PPI-weighted protein set distance was calculated between all significant DEPs (FDR < 0.1, pink diamond). To determine if this distance was smaller than would be expected by chance, size-matched sets were randomly chosen from the proteins detected in each experiment, and distance between these random sets was calculated 1000 times per pair. *P*-values were assigned by counting the fraction of times this random distance was less than the actual distance value between differential sets. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and middle hinge is the median (50th). Both whiskers extends from the hinge to the largest or smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). **i**, Protein set distances between the top 50 DEPs per mutation. For each pair of mutations, a PPI-weighted protein set distance was calculated as in **h**. Number of organoids used for the analyses are summarized in the Methods under “Statistics and reproducibility”. DEPs: differentially expressed proteins. MS: mass spectrometry.

[Extended Data Fig. 13 Mutations in ASD risk genes in human brain organoids converge on asynchronous development of shared neuronal classes.](#)

Conceptual schematics highlighting main results.

Supplementary information

[Supplementary Information](#)

Supplementary Notes and Supplementary References, and legends for Supplementary Fig. 1, Supplementary Tables 1–10 and Supplementary Video 1.

Reporting Summary

Supplementary Fig. 1

Uncropped gel images of western blots shown in this study.

Supplementary Table 1

Supplementary Table 2

Supplementary Table 3

Supplementary Table 4

Supplementary Table 5

Supplementary Table 6

Supplementary Table 7

Supplementary Table 8

Supplementary Table 9

Supplementary Table 10

Supplementary Video 1

Network bursting in a human brain organoid.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Paulsen, B., Velasco, S., Kedaigle, A.J. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022). <https://doi.org/10.1038/s41586-021-04358-6>

- Received: 02 November 2020
- Accepted: 14 December 2021
- Published: 02 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04358-6>

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/s41586-021-04358-6>

- Article
- [Published: 26 January 2022](#)

Cortical preparatory activity indexes learned motor memories

- [Xulu Sun](#) [ORCID: orcid.org/0000-0003-3076-0994](#)^{1,2 na1},
- [Daniel J. O'Shea](#) [ORCID: orcid.org/0000-0002-1366-1743](#)^{2,3 na1},
- [Matthew D. Golub](#) [ORCID: orcid.org/0000-0003-4508-0537](#)^{2,3},
- [Eric M. Trautmann](#)^{2,3},
- [Saurabh Vyas](#)^{2,4},
- [Stephen I. Ryu](#)^{3,5,6} &
- [Krishna V. Shenoy](#) [ORCID: orcid.org/0000-0003-1534-9240](#)^{2,3,4,6,7,8}

[Nature](#) volume **602**, pages 274–279 (2022)

- 5073 Accesses
- 73 Altmetric
- [Metrics details](#)

Subjects

- [Dynamical systems](#)
- [Learning and memory](#)
- [Motor cortex](#)

Abstract

The brain's remarkable ability to learn and execute various motor behaviours harnesses the capacity of neural populations to generate a variety of activity patterns. Here we explore systematic changes in preparatory activity in motor cortex that accompany motor learning. We trained rhesus monkeys to learn an arm-reaching task¹ in a curl force field that elicited new muscle forces for some, but not all, movement directions^{2,3}. We found that in a neural subspace predictive of hand forces, changes in preparatory activity tracked the learned behavioural modifications and reassociated⁴ existing activity patterns with updated movements. Along a neural population dimension orthogonal to the force-predictive subspace, we discovered that preparatory activity shifted uniformly for all movement directions, including those unaltered by learning. During a washout period when the curl field was removed, preparatory activity gradually reverted in the force-predictive subspace, but the uniform shift persisted. These persistent preparatory activity patterns may retain a motor memory of the learned field^{5,6} and support accelerated relearning of the same curl field. When a set of distinct curl fields was learned in sequence, we observed a corresponding set of field-specific uniform shifts which separated the associated motor memories in the neural state space^{7,8,9}. The precise geometry of these uniform shifts in preparatory activity could serve to index motor memories, facilitating the acquisition, retention and retrieval of a broad motor repertoire.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

[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.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Task design and behavioural performance.



Fig. 2: Changes in preparatory neural states accompanying learning.

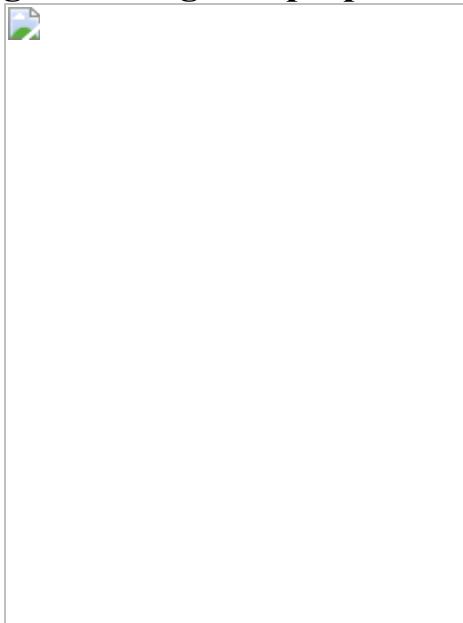


Fig. 3: Field-specific geometry of preparatory uniform shifts.

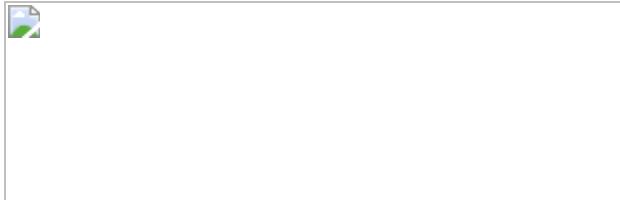


Fig. 4: Interference during learning of opposing curl fields at two targets 30° apart.

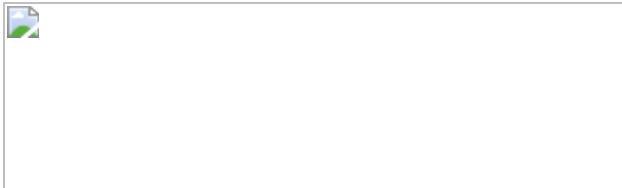
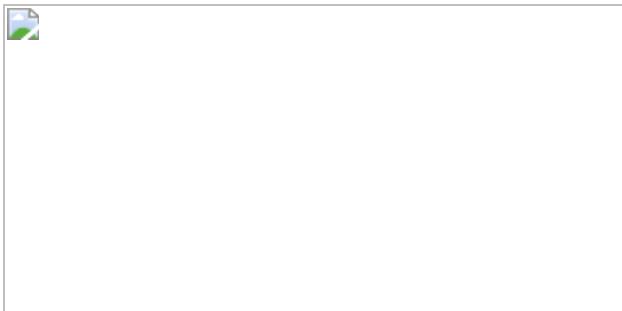


Fig. 5: Uniform-shift progression correlates with motor memory retention.



Data availability

The data that support the findings of the current study are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

Code availability

The code for the repertoire change analysis is available on github (https://github.com/mattgolub/bci_learning). The code for the TDR and uniform-shift analyses is available on github (<https://github.com/xlsun79/TDRandUniformShiftAnalyses>). All other code is available from the corresponding author upon reasonable request.

References

1. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
2. Shadmehr, R. Generalization as a behavioral window to the neural mechanisms of learning internal models. *Hum. Mov. Sci.* **23**, 543–568

(2004).

3. Howard, I. S. & Franklin, D. W. Neural tuning functions underlie both generalization and interference. *PLoS ONE* **10**, e0131268 (2015).
4. Golub, M. D. et al. Learning by neural reassociation. *Nat. Neurosci.* **21**, 607–616 (2018).
5. Li, C. S., Padoa-Schioppa, C. & Bizzi, E. Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron* **30**, 593–607 (2001).
6. Xiao, J. Premotor neuronal plasticity in monkeys adapting to a new dynamic environment. *Eur. J. Neurosci.* **22**, 3266–3280 (2005).
7. Sheahan, H. R., Franklin, D. W. & Wolpert, D. M. Motor planning, not execution, separates motor memories. *Neuron* **92**, 773–779 (2016).
8. Sheahan, H. R., Ingram, J. N., Žalalytė, G. M. & Wolpert, D. M. Imagery of movements immediately following performance allows learning of motor skills that interfere. *Sci. Rep.* **8**, 14330 (2018).
9. Schween, R., Langsdorf, L., Taylor, J. A. & Hegele, M. How different effectors and action effects modulate the formation of separate motor memories. *Sci. Rep.* **9**, 17040 (2019).
10. Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L. & Haith, A. M. Motor learning. *Compr. Physiol.* **9**, 613–663 (2019).
11. Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
12. Herzfeld, D. J., Vaswani, P. A., Marko, M. K. & Shadmehr, R. A memory of errors in sensorimotor learning. *Science* **345**, 1349–1353 (2014).

13. Howard, I. S., Wolpert, D. M. & Franklin, D. W. The effect of contextual cues on the encoding of motor memories. *J. Neurophysiol.* **109**, 2632–2644 (2013).
14. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
15. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
16. Sauerbrei, B. A. et al. Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
17. Stavisky, S. D., Kao, J. C., Ryu, S. I. & Shenoy, K. V. Trial-by-trial motor cortical correlates of a rapidly adapting visuomotor internal model. *J. Neurosci.* **37**, 1721–1732 (2017).
18. Vyas, S. et al. Neural population dynamics underlying motor learning transfer. *Neuron* **97**, 11771186.e3 (2018).
19. Perich, M. G., Gallego, J. A. & Miller, L. E. A neural population mechanism for rapid learning. *Neuron* **100**, 964–976.e7 (2018).
20. Vyas, S., O’Shea, D. J., Ryu, S. I. & Shenoy, K. V. Causal role of motor preparation during error-driven learning. *Neuron* **106**, 329–339.e4 (2020).
21. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
22. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
23. Even-Chen, N., Sheffer, B., Vyas, S., Ryu, S. I. & Shenoy, K. V. Structure and variability of delay activity in premotor cortex. *PLoS*

Comput. Biol. **15**, e1006808 (2019).

24. Arce, F., Novick, I., Mandelblat-Cerf, Y. & Vaadia, E. Neuronal correlates of memory formation in motor cortex after adaptation to force field. *J. Neurosci.* **30**, 9189–9198 (2010).
25. 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).
26. Jarosiewicz, B. et al. Functional network reorganization during learning in a brain-computer interface paradigm. *Proc. Natl Acad. Sci. USA* **105**, 19486–19491 (2008).
27. Chase, S. M., Kass, R. E. & Schwartz, A. B. Behavioral and neural correlates of visuomotor adaptation observed through a brain–computer interface in primary motor cortex. *J. Neurophysiol.* **108**, 624–644 (2012).
28. Oby, E. R. et al. New neural activity patterns emerge with long-term learning. *Proc. Natl Acad. Sci. USA* **116**, 15210–15215 (2019).
29. Sakellaridi, S. et al. Intrinsic variable learning for brain–machine interface control by human anterior intraparietal cortex. *Neuron* **102**, 694–705.e3 (2019).
30. Hwang, E. J., Bailey, P. M. & Andersen, R. A. Volitional control of neural activity relies on the natural motor repertoire. *Curr. Biol.* **23**, 353–361 (2013).
31. Rokni, U., Richardson, A. G., Bizzi, E. & Seung, H. S. Motor learning with unstable neural representations. *Neuron* **54**, 653–666 (2007).
32. Franklin, D. W., Osu, R., Burdet, E., Kawato, M. & Milner, T. E. Adaptation to stable and unstable dynamics achieved by combined impedance control and inverse dynamics model. *J. Neurophysiol.* **90**, 3270–3282 (2003).

33. Heald, J. B., Franklin, D. W. & Wolpert, D. M. Increasing muscle co-contraction speeds up internal model acquisition during dynamic motor learning. *Sci. Rep.* **8**, 16355 (2018).
34. Richardson, A. G., Borghi, T. & Bizzi, E. Activity of the same motor cortex neurons during repeated experience with perturbed movement dynamics. *J. Neurophysiol.* **107**, 3144–3154 (2012).
35. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* **68**, 387–400 (2010).
36. Russo, A. A. et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953–966.e8 (2018).
37. Ames, K. C., Ryu, S. I. & Shenoy, K. V. Neural dynamics of reaching following incorrect or absent motor preparation. *Neuron* **81**, 438–451 (2014).

Acknowledgements

We thank the members of the Shenoy laboratory at Stanford University for comments and discussions on the methods and results; T. Fisher for help with experiments; M. Risch, M. Wechsler and R. Reeder for expert veterinary care; B. Davis for administrative support; and W. L. Gore for donating Preclude artificial dura used as part of the chronic electrode array implantation procedure. X.S. was supported by a Stanford Interdisciplinary Graduate Fellowship, a Stanford Bio-X Honorary Fellowship and Stanford Department of Biology Funding. D.J.O. was supported by a US National Science Foundation graduate research fellowship and a Stanford Graduate Fellowship. M.D.G. was supported by an NIH K99/R00 award NIMH-K99MH121533. E.M.T. was supported by an NIH NRSA grant 1F31NS089376-01, a Stanford Graduate Fellowship and an NSF IGERT grant 0734683. S.V. was supported by an NIH F31 Ruth L. Kirschstein National Research Service Award 1F31NS103409-01, an NSF Graduate Research Fellowship and a Ric Weiland Stanford Graduate Fellowship.

K.V.S. was supported by the following awards: NIH Director's Pioneer Award 8DP1HD075623, Defense Advanced Research Projects Agency (DARPA) Biological Technology Office (BTO) 'NeuroFAST' award W911NF-14-2-0013, the Simons Foundation Collaboration on the Global Brain awards 543045, the Office of Naval Research N000141812158 and the Howard Hughes Medical Institute.

Author information

Author notes

1. These authors contributed equally: Xulu Sun, Daniel J. O'Shea

Affiliations

1. Department of Biology, Stanford University, Stanford, CA, USA

Xulu Sun

2. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA

Xulu Sun, Daniel J. O'Shea, Matthew D. Golub, Eric M. Trautmann, Saurabh Vyas & Krishna V. Shenoy

3. Department of Electrical Engineering, Stanford University, Stanford, CA, USA

Daniel J. O'Shea, Matthew D. Golub, Eric M. Trautmann, Stephen I. Ryu & Krishna V. Shenoy

4. Department of Bioengineering, Stanford University, Stanford, CA, USA

Saurabh Vyas & Krishna V. Shenoy

5. Department of Neurosurgery, Palo Alto Medical Foundation, Palo Alto, CA, USA

Stephen I. Ryu

6. Department of Neurosurgery, Stanford University, Stanford, CA, USA

Stephen I. Ryu & Krishna V. Shenoy

7. Department of Neurobiology, Stanford University, Stanford, CA, USA

Krishna V. Shenoy

8. Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

Krishna V. Shenoy

Contributions

X.S. conceived the project and designed the experiments with input from D.J.O. and M.D.G. X.S. conducted the V-probe, Utah array and EMG recordings. X.S. and D.J.O. conducted the Neuropixels recordings, with extensive assistance from E.M.T. S.V. collected the VMR data. X.S. performed data analysis with significant assistance from D.J.O. and M.D.G., as well as input from S.V. S.I.R. led the recording chamber and array implantation surgeries. X.S. and D.J.O. wrote the manuscript with input and editing from all authors. K.V.S. was involved with all aspects of the research.

Corresponding authors

Correspondence to [Xulu Sun](#) or [Krishna V. Shenoy](#).

Ethics declarations

Competing interests

K.V.S. is a consultant to Neuralink Corp. and CTRL-Labs Inc., in the Reality Labs Division of Meta (formerly FaceBook). He is on the Scientific

Advisory Boards of Inscopix Inc., Mind X Inc. and Heal Inc. These entities did not support or influence this work. The authors declare no other competing interests.

Peer review information

Nature thanks Aaron Batista, J. Andrew Pruszynski and the other, anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

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 information for task design, recording sites and behavioural performance.](#)

a, Spatial arrangements of the 12 reaching targets. The target density near the trained target (down, up or right) was higher in order to sample more neural states for reaches that were more likely to be altered by learning, for the purpose of studying generalization of learning³. Ideally, one would like to have equally-spaced reach targets as dense as possible, but because monkeys could perform a limited number of trials each day, a compromise solution was to increase the density of targets near the trained target. Note that the curl field can be either clockwise (CW) or counterclockwise (CCW). Here we show CW fields as an example. **b**, Utah-array implant locations in monkey U and recording sites in monkey V. Recordings were performed in PMd and M1 in the hemisphere contralateral to the reaching arm. Top panel: an intraoperative photo of three Utah-array implants in monkey U. Bottom panel: a schematic map illustrating the approximate locations of recording sites in monkey V based on stereotactic coordinates; data in this work included units recorded from multi-electrode V-probes and

Neuropixels probes. Histology has not yet been done on either monkey. Using the cortical landmarks, we estimated that the recording sites in monkey V largely overlapped with the lateral half of the area covered by the three Utah arrays in monkey U. AS: Spur of arcuate. CS: central sulcus. PCD: precentral dimple. **c**, We computed behavioural (top panel) and neural generalization (bottom panel) with the sign of the effects flipped for CCW fields to match the effects of CW fields, compared to Fig. 1d and Fig. 2c. We found a spatial asymmetry in behavioural and neural generalization, with more learning in the ‘push’ direction (i.e., the direction to oppose the curl field). Error bars, s.e.m. across sessions (monkey U, n = 4, 3; monkey V, n = 5, 3). **d**, Top panel: trial-averaged hand speed in different blocks over multiple learning sessions. Shaded area, s.e.m. across sessions (monkey U, n = 4; monkey V, n = 5). Bottom panel: compensatory hand force perpendicular to the reach direction in one example session. Hand force in late-learning trials (dark red) showed a more stereotypical, less variable temporal pattern with an earlier onset than in early-learning trials (light red). Time zero, movement onset.

[Source data](#)

Extended Data Fig. 2 Additional results of neural activity patterns during curl field learning and VMR learning.

a, Initial hand forces predicted by the 2D TDR preparatory states were correlated with real forces of the upcoming movement (the slope is 0.41 and the intercept is 0.02 with $\sqrt{(\{\{ \rm{R} \}\})^2} = 0.59$ and $P = 6.06 \times 10^{-9}$). The sign of hand force indicates its direction. Lighter dots, earlier learning trials; darker dots, later learning trials. **b**, Single-trial prediction MSE of initial hand forces was significantly smaller using original data than shuffled data (two-sided Wilcoxon rank-sum test: $P = 0.0006$ for both training and test sets of monkey U and $P = 0.008$ for both training and test sets of monkey V). Training set: before-learning trials. Test set: learning trials. Control results (blue) were forces predicted by models built from training sets that had neural and behavioural data shuffled. One datapoint per session. **c**, Changes of preparatory states in the force-predictive TDR subspace reflected generalization of learning, quantified as the rotatory angle from before-learning to error-clamp neural states. Zero degree on the

x axis, the trained target. Error bars, s.e.m. across sessions ($n = 5, 3$). **d**, Normalized single-trial neural shift during learning along the uniform-shift learning axis. Solid line: linear-log regression ($n = 1200, 900$). **e**, Preparatory neural repertoires changed similarly for trained and untrained reaches. Black: no-learning control results ($n = 36$). Blue: far targets more than 45 degrees from the trained target ($n = 15$). Red: near targets within 45 degrees from the trained target ($n = 21$). One-sided Wilcoxon rank-sum test: $P_{black \ vs. \ blue} = 2.33 \times 10^{-7}$, $P_{black \ vs. \ red} = 4.74 \times 10^{-8}$, $P_{blue \ vs. \ red} = 0.059$. **f, g**, VMR learning results. **f**, Preparatory neural states projected to PCs 1-3. After-learning states (diamonds) were mixed with before-learning states (circles). One example session. **g**, Preparatory and peri-movement neural activity patterns did not show repertoire change during VMR learning. One-sided Wilcoxon rank-sum test: $P > 0.1$ for all comparisons. Three learning sessions ($n = 24$) and three control sessions ($n = 24$) for both monkeys. **h, i**, Neural preparatory states in the 3D TDR subspace. The 3D subspace was constructed by TDR capturing the variance due to initial hand forces and a binary indicator of trial conditions (an indicator of before-learning versus after-learning). One example session. **h**, In the force-predictive TDR subspace, rotatory shifts of preparatory neural states were similar to Fig. 2b. **i**, Along the TDR 3 axis (the binary indicator axis), this 3D model revealed a uniform shift similar to what we observed along PC 3 in the PCA subspace (Fig. 2d). For all the box plots, the central line indicates the median, the bottom and top edges indicate the 25th and 75th percentiles of the data, and the whiskers extend to the 5th and 95th percentiles of the data.

Source data

Extended Data Fig. 3 Relationships between neural population dimensions and total neural variance explained by different dimensions.

a, Pairwise dot products between neural population dimensions. Values close to 1 indicate that two dimensions are closely aligned and values close to 0 indicate that two dimensions are nearly orthogonal. In each session, we calculated the dot product of TDR 1 and PC 1 and the dot product of TDR 1 and PC 2, and took the larger value of the two dot products (TDR 1 vs. PC

1 / 2). We then calculated the dot product of TDR 2 and the PC axis not used for multiplying with TDR 1 (TDR 2 vs. PC 1 / 2). The PC 1 / 2 plane largely overlapped with the TDR 1 / 2 plane (black). The TDR 1, TDR 2, and hand-speed TDR axes were all nearly orthogonal to the uniform-shift learning axis (blue and red). PC 3 largely overlapped with the uniform-shift learning axis (yellow). Two-sided Wilcoxon rank-sum test: monkey U, $**P = 4.04 \times 10^{-3}$, $*P = 0.029$; monkey V, $***P = 6.66 \times 10^{-4}$, $**P = 7.94 \times 10^{-3}$. Right panel: a schematic illustration of projecting data points from axis $\backslash(\alpha\backslash)$ to axis $\backslash(\beta\backslash)$ and the corresponding dot product. **b, c**, The portion of total neural activity variance explained by the TDR 1 and TDR 2 (hand force) axes, hand-speed TDR axis, uniform-shift axis and PCs 1-3. **b**, Two-sided Wilcoxon rank-sum test: monkey U, $*P = 0.029$, n.s. $P = 1$; monkey V, $*P = 0.016$ and 0.032 , n.s. $P = 0.42$. **a-c**, $n = 4$ (monkey U) and $n = 5$ (monkey V). For all the box plots, the central line indicates the median, the bottom and top edges indicate the 25th and 75th percentiles of the data, and the whiskers extend to the 5th and 95th percentiles of the data.

[Source data](#)

[Extended Data Fig. 4 EMG signals of 6 upper limb muscles \(bicep, radialis flexor, radialis extensor, pectoralis, posterior deltoid, lateral deltoid\).](#)

Time zero, movement onset. One example condition (CW curl field applied to down reaches). Shaded area, s.e.m. across trials. **a**, EMG signals in before-learning, late-learning, error-clamp and late-washout blocks. Muscle activity did not show signs of muscle co-contraction during learning (red). Muscle activity during the preparatory period remained flat and around the same level across all blocks (two-sided rank-sum test: $P < 0.0001$ for comparing late-learning or error-clamp bicep activity with before-learning or late-washout bicep activity during the preparatory period; $P > 0.3$ for all the other pairs of comparison during the preparatory period). Muscle activity patterns in before-learning (black) and late-washout trials (blue) were very similar. Muscle activity patterns in late-learning (red) and error-clamp trials (purple) were very similar. **b**, EMG signals in before-learning (black) and error-clamp (purple) blocks did not show a uniform shift across all 12 reaching targets. For all six muscles, EMG activity after learning

increased in some directions and decreased in other directions. Muscle activity of reaching to the target 135 degrees away from the trained target (i.e., far targets with almost no behavioural generalization, see Fig. [1c](#)) in before-learning and error-clamp trials showed similar temporal patterns.

[Source data](#)

[**Extended Data Fig. 5 The uniform shift is not due to the error clamp.**](#)

a, Preparatory states in error-clamp (diamonds) and no-clamp (circles) trials were not significantly different in no-learning control sessions (Hotelling's T² test: $P > 0.5$ for all control sessions of monkeys U and V). One example session is shown. **b**, In the error-clamp block (block iii) of learning sessions, the late-learning preparatory state and error-clamp state of the trained target were not significantly different (Hotelling's T² test: $P > 0.05$ for all learning sessions of monkey U and $P > 0.1$ for all learning sessions of monkey V). Arrows point to the before-learning state (purple circle) and error-clamp state (purple diamond) of the trained target. One example session is shown. **c**, Late-learning preparatory states comprised a new neural repertoire following learning but not during control reaches with random force perturbations. One-sided Wilcoxon rank-sum test: *** $P = 1.83 \times 10^{-5}$, n.s. $P = 0.99$ ($n = 12$ per box). The results are similar to Fig. [2g](#) where we used after-learning, error-clamp trials to compute the repertoire change. **d–f**, The uniform shift was not due to reorientation of the TDR plane in error-clamp trials. **d**, TDR axes using only error-clamp trials (TDR-EC) and TDR axes using only before-learning trials (TDR-BL) were largely aligned. The error-clamp TDR 1 axis and before-learning TDR 1 axis were highly aligned, and same for TDR 2 axes (black). The error-clamp TDR axes were nearly orthogonal to the uniform-shift axis (red), similar to the before-learning TDR axes shown in Extended Data Fig. [3a](#). One-sided Wilcoxon rank-sum test: monkey U, $n = 4$ and * $P = 0.014$; monkey V, $n = 5$ and ** $P = 4.0 \times 10^{-3}$. **e**, In the interference experiments, the force-predictive TDR planes constructed from only before-learning (BL) trials, only error-clamp (EC) trials, and both (BL+EC) were significantly aligned. One-sided signed-rank test comparing dot products to 0.8: $n = 8$ from

monkeys U + V, $**P = 3.9 \times 10^{-3}$. **f**, In the interference experiments, preparatory neural states showed similar patterns in the force-predictive TDR plane built from only before-learning (BL) trials, only error-clamp (EC) trials, or both (BL+EC). One example session. Small black arrows point to the before-learning states of trained targets. For all the box plots, the central line indicates the median, the bottom and top edges indicate the 25th and 75th percentiles of the data, and the whiskers extend to the 5th and 95th percentiles of the data.

[Source data](#)

Extended Data Fig. 6 Stability of multi-session recordings and the geometric relationship between uniform-shift axes for learning multiple curl fields.

a, b, Spike waveforms and peristimulus time histograms (PSTHs) of three example neurons across sessions. The same 71 neurons from monkey U Utah-array recordings were selected post-hoc by comparing waveform correlations and tracked over five successive sessions. **a**, All selected neurons had nearly-identical waveforms. **b**, Like the three example neurons, most selected neurons had similar direction-tuning properties for before-learning reaches across sessions. Go, go cue. Move, movement onset. **c**, Distribution of dot products between uniform-shift axes for learning two opposite fields applied at one reach target sequentially (green) or two distinct fields applied at different reach targets sequentially (purple). We compared them with simulated distributions of dot products between uniform shifts predicted by ‘orthogonal’ (red, around 0), ‘parallel’ (blue, around 1) and ‘antiparallel’ (black, around -1) relationships (see Measurement of geometric relationships between uniform-shift axes in Methods). Top inset, the zoom-in view of each simulated distribution.

[Source data](#)

Extended Data Fig. 7 Interference and sequential learning of opposite curl fields applied at two targets 30 degrees apart, monkey V results.

a, Behavioural learning quantified by lateral hand deviation. Lateral hand deviation in both curl fields decreased slowly during simultaneous learning (blue) and further reduced during sequential learning (red). One-sided Wilcoxon rank-sum test: $P_{CCW} = 1.30 \times 10^{-10}$, $P_{CW} = 1.51 \times 10^{-11}$. **b**, Behavioural and neural generalization of simultaneously learning two fields. Behavioural generalization was measured by perpendicular hand force differences between error-clamp and before-learning trials, and neural generalization was measured by the rotatory angle between before-learning and error-clamp neural states. Zero degree on the x axis, the middle target between the two trained targets. Error bars, s.e.m. from resampling (100 repeats). **c**, Preparatory neural states in the force-predictive TDR subspace. Before-learning states (circles) spatially organized corresponding to reach directions. Error-clamp states (diamonds) rotated counterclockwise for reach targets near the CW field and clockwise for targets near the CCW field. Preparatory states of the two trained targets (triangles) rotated opposite their curl field directions in blocks ii and iii (left panel), and further separated in the sequential-learning block (right panel). Small yellow arrows point to the before-learning states of trained targets. Neural states of seven nearest targets are visualized in c, and quantified neural changes for all 11 targets are shown in b. **d, Left panel:** preparatory activity projected into the subspace spanned by the two field-specific uniform shifts. Without orthogonalization, these two uniform shifts were 128 degrees apart. The uniform shifts were orthonormalized before projection such that: x axis = field 1 uniform shift, y axis = field 2 uniform shift – the projection of field 2 uniform shift on field 1 uniform shift. During simultaneous learning (blocks ii and iii), preparatory states of each field (orange and pink) moved in its specific uniform-shift direction while also progressing in the other uniform-shift direction, and were significantly separated (Hotelling's T^2 test: $P = 2.58 \times 10^{-6}$). Grey arrows illustrate the hypothesized trial-by-trial progression of preparatory states for both fields during simultaneous learning. During sequential learning, preparatory states of each field (green and purple) further separated (Hotelling's T^2 test: $P = 0$). **Right panel:** A residual interference shift orthogonal to the field-specific uniform shifts occurred during simultaneous learning. Neural states in the left panel are the projection of neural states into the grey plane in the right panel.

[Source data](#)

Extended Data Fig. 8 Interference and sequential learning of opposite curl fields applied to the same target.

a, Block design of the interference experiment. Same as in Fig. 4a except that two opposite curl fields were applied to the same reach target. **b**, Behavioural learning quantified by lateral hand deviation. Hand lateral deviation in both curl fields only slightly decreased during simultaneous learning (blue), and significantly reduced during sequential learning (red). One-sided rank-sum test: monkey U, *** $P = 1.51 \times 10^{-11}$; monkey V, *** $P = 1.51 \times 10^{-11}$. **c**, Preparatory neural states in the force-predictive TDR subspace. In blocks ii and iii (left panel), preparatory states of the two curl fields (triangles) were mixed together around the before-learning state (circle). Error-clamp states (diamonds) of most targets shifted from their corresponding before-learning states. These shifts did not show coherent patterns across targets or monkeys and were likely unrelated to learning. In the sequential-learning block (right panel), preparatory states of the two curl fields (triangles) gradually rotated opposite their curl field directions. The small arrow points to the before-learning state of the reach target that later had curl fields (trained target). Neural states for seven nearest targets are shown. **d**, Perpendicular hand force differences between error-clamp and before-learning trials (top panel), and the rotatory angle from before-learning to error-clamp neural states (bottom panel), did not show coherent patterns across targets or monkeys. Zero degree on the x axis, the trained target. Error bars, s.e.m. from resampling (100 repeats). **e**, Uniform shifts for learning two curl fields and the residual interference shift were defined in the same way as in Fig. 4e,f. The two field-specific uniform shifts were close to antiparallel (monkey U, dot product = - 0.79; monkey V, dot product = -0.64), and so we could visualize preparatory neural states in a 2D plane spanned by the field 1 uniform shift and the residual interference shift. In blocks ii and iii, preparatory states of the two curl fields (orange and pink) shifted away from the before-learning centroid (grey circle) along the residual interference axis, but they remained close to each other (Hotelling's T^2 test: monkey U, $P = 0.66$; monkey V, $P = 0.98$). During sequential learning, preparatory states of the two curl fields (green and purple) were separated by opposite uniform shifts (Hotelling's T^2 test: monkey U, $P = 2.49 \times 10^{-4}$; monkey V, $P = 2.90 \times 10^{-5}$). **b–e**, One session

per monkey. Though just one session of this interference experiment was performed with each monkey, the results were consistent across monkeys and complimentary to findings when monkeys learned multiple fields sequentially, which supported the indexing hypothesis (Fig. 3).

Source data

Extended Data Fig. 9 Monkey V washout results and additional information on the neural population correlate of motor memory retention.

a, The angular difference between washout states and the before-learning state gradually decreased on a single-trial basis (grey dots: single-trial data points from all sessions; solid line: linear regression). Normalized against the maximum in each session. **b**, Distance between washout states and the before-learning state along the uniform-shift learning axis did not show a significant trend of increase or decrease (solid line: linear regression). Each dot is a single trial. **c**, Hand deviation was smaller during relearning than during initial learning (one-sided Wilcoxon rank-sum test: $P = 0.0015$). **d**, Neural trajectories approached late-learning trajectories faster during relearning than initial learning. One-sided rank-sum test: $P = 6.18 \times 10^{-4}$. **e**, Shaded area, s.e.m. across sessions ($n = 3$). **f**, Distribution of dot products between uniform-shift learning and washout axes (purple), compared to simulated distributions of dot products between uniform-shift axes predicted by orthogonal (red), parallel (blue) and antiparallel (black) relationships. **g**, Distribution of dot products between uniform-shift axes for learning the same curl field in two sessions 18 days apart (green, close to 1). **h**, Neural trajectories of before-learning, late-learning, late-washout and late-relearning conditions (-150 to $+150$ ms from target onset, covered by the grey circle; -50 to $+50$ ms from the go cue, covered by the grey ellipse; and -200 to $+400$ ms from movement onset). Movement preparation and execution periods are noted on the trajectories. Black arrows show the direction of neural trajectories. The late-washout trajectory (green) was less similar to the before-learning trajectory (black) during movement preparation than execution. TO: target onset. GC: go cue. One example session. **i**, During preparatory period (prep), the similarity between late-washout and before-learning neural trajectories was significantly lower than

the similarity between before-learning neural trajectories. During movement period (move), the similarity between late-washout and before-learning neural trajectories could compare to the similarity between before-learning neural trajectories. One-sided rank-sum test: seven sessions from monkeys U + V, $**P = 0.0012$, n.s. $P = 0.50$. For each box, the central line indicates the median, the bottom and top edges indicate the 25th and 75th percentiles of the data, and the whiskers extend to the 5th and 95th percentiles of the data.

[Source data](#)

[Extended Data Fig. 10 Neural population activity patterns in other time windows.](#)

a, PCs 1-4 during -100 to +100 ms from target onset (TO), -100 to +100 ms from go cue (GC), and -200 to +400 ms from movement onset (MOO). We applied PCA to trial-averaged neural activity for different reach directions in the before-learning and error-clamp blocks. Across all learning sessions in both monkeys, PC 1 explained 30 – 40% of the total variance, PC 2 explained 10 - 20%, PC 3 explained 8 - 10%, and PC 4 explained 6 - 8%. Neural trajectories in PCs 1-4 were bundled together around the target onset time window and diverged around the go cue time window (preparatory period). Error-clamp neural trajectories all shifted from their corresponding before-learning trajectories in PC 3. The time window -50 to +50 ms from go cue (grey shadow) we chose for preparatory neural state analysis was within the preparatory period and had stronger neural tuning than earlier time windows (e.g., the first 100 ms after target onset). Top left inset: color-coded reach directions. **b**, No significant neural repertoire change of baseline neural activity (before target onset on each trial) after learning the curl field. Black: no-learning control results ($n = 36, 36$). Red: learning results ($n = 48, 36$). One-sided Wilcoxon rank-sum test: monkey U, $P = 0.999$; monkey V, $P = 0.595$. **c–i**, Patterns of peri-movement neural population states. **c**, Peri-movement states of before-learning (color circles), learning (grey circles) and after-learning error-clamp (diamonds) reaches projected to PCs 1-3. After-learning states of the trained target and its nearby untrained targets left the before-learning states. One example session. **d**, Quantification of neural shift during learning along the ‘peri-

movement shift axis' that connected the before-learning and error-clamp states of the trained target, normalized against the distance between these two states. $n = 4$ (monkey U), 5 (monkey V). Cuzick's test: monkey U, $P = 0.032$; monkey V, $P = 3.92 \times 10^{-5}$. **e**, The Euclidian distance between before-learning and after-learning peri-movement states showed bell-shaped local generalization. Error bars, s.e.m. across sessions (monkey U: $n = 4, 3$; monkey V: $n = 5, 3$). **f**, Peri-movement activity patterns showed significantly greater repertoire change for the trained target and near targets than far targets (monkey U: $n = 28, 20$; monkey V: $n = 21, 15$). Black: no-learning control sessions ($n = 36$ for both monkeys). One-sided Wilcoxon rank-sum test: monkey U, $P_{black \ vs. \ blue} = 0.26$, $P_{black \ vs. \ red} = 4.52 \times 10^{-6}$, $P_{blue \ vs. \ red} = 0.002$; monkey V, $P_{black \ vs. \ blue} = 3.70 \times 10^{-7}$, $P_{black \ vs. \ red} = 6.02 \times 10^{-9}$, $P_{blue \ vs. \ red} = 5.29 \times 10^{-4}$. **g**, Peri-movement states in the same PCA subspace during washout. **h**, Distance between washout and before-learning states decreased significantly along the peri-movement shift axis. Normalized against the distance between the before-learning and after-learning states of the trained target. $n = 4$ (monkey U), 5 (monkey V). Cuzick's test: monkey U, $P = 0.0077$; monkey V, $P = 0.0028$. **i**, Pairwise dot products between peri-movement neural dimensions. PCs 1-3 significantly overlapped with the peri-movement shift (one-sided signed-rank test compared to 0: monkeys U + V, $n = 9$ and $P = 0.002$ for all comparisons). TDR 1 / 2 axes also significantly overlapped with the peri-movement shift (one-sided signed-rank test compared to 0: monkeys U + V, $n = 18$ and $P = 1.07 \times 10^{-4}$). For all the box plots, the central line indicates the median, the bottom and top edges indicate the 25th and 75th percentiles of the data, and the whiskers extend to the 5th and 95th percentiles of the data.

[Source data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Notes 1–3, the Methods and discussion. Supplementary Note 1 contains a brief comparison of changes of initial conditions versus changes of neural dynamics matrix, from the linear

dynamical systems perspective. Supplementary Note 2 includes an extensive set of control analyses and results to support the idea that the observed uniform shift did not directly mirror movement output and was learning-related. Supplementary Note 3 contains additional neural results during the peri-movement period that differed from the preparatory neural results. The discussion includes a more detailed discussion of the results.

Reporting Summary

Supplementary Table 1

Detailed statistics for main figures.

Supplementary Table 2

Variance of the uniform shift explained by multiple alternative models, as discussed in Supplementary Note 2.

Peer Review File

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. 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

Sun, X., O’Shea, D.J., Golub, M.D. *et al.* Cortical preparatory activity indexes learned motor memories. *Nature* **602**, 274–279 (2022).
<https://doi.org/10.1038/s41586-021-04329-x>

- Received: 25 January 2020
- Accepted: 09 December 2021
- Published: 26 January 2022
- Issue Date: 10 February 2022

- DOI: <https://doi.org/10.1038/s41586-021-04329-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/s41586-021-04329-x>

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

- Article
- [Published: 22 December 2021](#)

Monocotyledonous plants graft at the embryonic root–shoot interface

- [Gregory Reeves](#) ORCID: [orcid.org/0000-0002-3474-8960^{1,2}](https://orcid.org/0000-0002-3474-8960),
- [Anoop Tripathi](#) ORCID: [orcid.org/0000-0003-1195-6609¹](https://orcid.org/0000-0003-1195-6609),
- [Pallavi Singh](#) ORCID: [orcid.org/0000-0003-3694-6378¹](https://orcid.org/0000-0003-3694-6378),
- [Maximillian R. W. Jones](#) ORCID: [orcid.org/0000-0002-4887-591X¹](https://orcid.org/0000-0002-4887-591X),
- [Amrit K. Nanda³](#),
- [Constance Musseau³](#),
- [Melanie Craze](#) ORCID: [orcid.org/0000-0003-0757-0914²](https://orcid.org/0000-0003-0757-0914),
- [Sarah Bowden²](#),
- [Joseph F. Walker⁴](#),
- [Alison R. Bentley](#) ORCID: [orcid.org/0000-0001-5519-4357^{2 nAff5}](https://orcid.org/0000-0001-5519-4357),
- [Charles W. Melnyk](#) ORCID: [orcid.org/0000-0003-3251-800X³](https://orcid.org/0000-0003-3251-800X) &
- [Julian M. Hibberd](#) ORCID: [orcid.org/0000-0003-0662-7958¹](https://orcid.org/0000-0003-0662-7958)

[Nature](#) volume **602**, pages 280–286 (2022)

- 10k Accesses
- 1 Citations
- 403 Altmetric
- [Metrics details](#)

Subjects

- [Molecular evolution](#)

- [Plant development](#)
- [Plant physiology](#)

Abstract

Grafting is possible in both animals and plants. Although in animals the process requires surgery and is often associated with rejection of non-self, in plants grafting is widespread, and has been used since antiquity for crop improvement¹. However, in the monocotyledons, which represent the second largest group of terrestrial plants and include many staple crops, the absence of vascular cambium is thought to preclude grafting². Here we show that the embryonic hypocotyl allows intra- and inter-specific grafting in all three monocotyledon groups: the commelinids, lilioids and alismatids. We show functional graft unions through histology, application of exogenous fluorescent dyes, complementation assays for movement of endogenous hormones, and growth of plants to maturity. Expression profiling identifies genes that unify the molecular response associated with grafting in monocotyledons and dicotyledons, but also gene families that have not previously been associated with tissue union. Fusion of susceptible wheat scions to oat rootstocks confers resistance to the soil-borne pathogen *Gaeumannomyces graminis*. Collectively, these data overturn the consensus that monocotyledons cannot form graft unions, and identify the hypocotyl (mesocotyl in grasses) as a meristematic tissue that allows this process. We conclude that graft compatibility is a shared ability among seed-bearing plants.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Mesocotyl allows monocotyledon grafting.

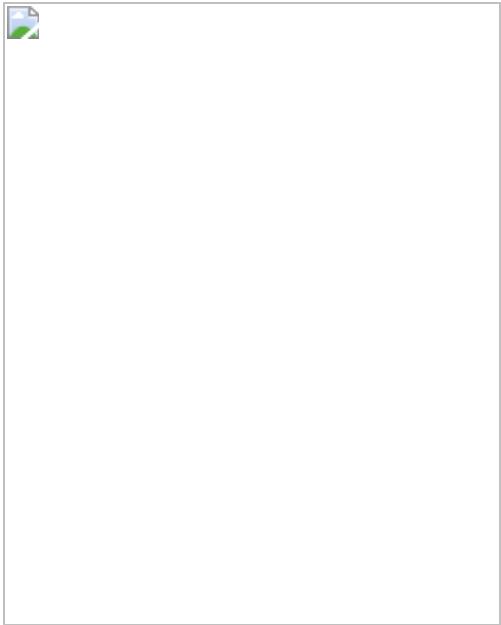


Fig. 2: Molecular analysis of monocotyledon graft unions.

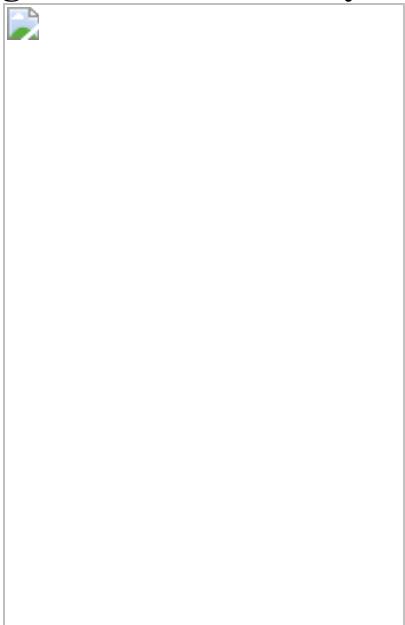


Fig. 3: Graft compatibility is widespread in the monocotyledons.

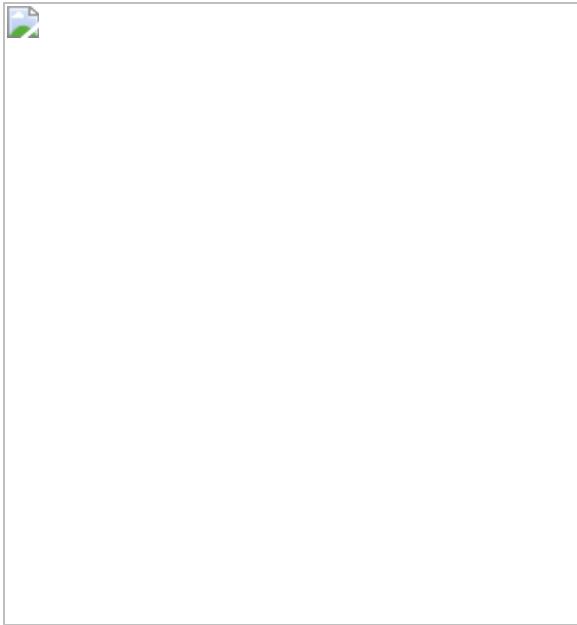
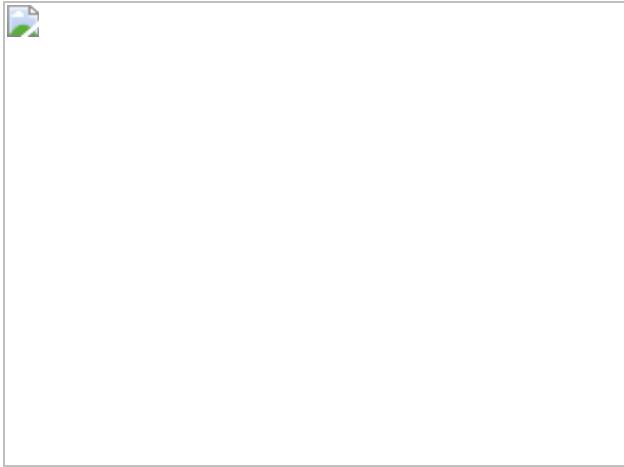


Fig. 4: Grafting monocotyledons modifies phenotypes.



Data availability

All data are available in the manuscript, Extended Data and supplementary materials. Deep sequencing reads for grafted, non-grafted and wounded rice have been deposited into the NCBI Sequence Read Archive (SRA) with accession number [PRJNA734117](#). Source data and custom coding scripts for plotting have been deposited into the GitHub repository: https://github.com/GregReeves/Reeves2021_MonocotGrafting. Source data are provided with this paper.

Code availability

Source data and custom R scripts used for individual plots and for browsing transcriptome results, as well as source data and Python scripts used for constructing monocotyledon phylogenies have been deposited at https://github.com/GregReeves/Reeves2021_MonocotGrafting. None of the custom code is central to conclusions of the manuscript.

References

1. Mudge, K., Janick, J., Scofield, S. & Goldschmidt, E. E. A history of grafting. *Hortic. Rev.* **35**, 437–493 (2009).
2. Melnyk, C. W. & Meyerowitz, E. M. Plant grafting. *Curr. Biol.* **25**, 183–188 (2015).
3. The Plant List v.1.1. <http://www.theplantlist.org/> (2013).
4. Calderini, I. M. Essai d’expériences sur la greffe des graminées [Experimental trials on grafting grasses]. *Ann. Sci. Nat. Bot.* **1846**, 131–133 (1846).
5. Muzik, T. J. & La Rue, C. D. The grafting of large monocotyledonous plants. *Science* **116**, 589–591 (1952).
6. Obolensky, G. Grafting of plant embryos and the use of ultrasonics. *Plant Food Hum. Nutr.* **7**, 273–288 (1960).
7. McCann, M. C. Chimeric plants—the best of both worlds. *Science* **369**, 618–619 (2020).
8. Melnyk, C. W., Schuster, C., Leyser, O. & Meyerowitz, E. M. A developmental framework for graft formation and vascular reconnection in *Arabidopsis thaliana*. *Curr. Biol.* **25**, 1306–1318 (2015).

9. Notaguchi, M. et al. Cell–cell adhesion in plant grafting is facilitated by β -1,4-glucanases. *Science* **369**, 698–702 (2020).
10. Melnyk, C. W. et al. Transcriptome dynamics at *Arabidopsis* graft junctions reveal an intertissue recognition mechanism that activates vascular regeneration. *Proc. Natl Acad. Sci. USA.* **115**, E2447–E2456 (2018).
11. Iwase, A. et al. The AP2/ERF transcription factor WIND1 controls cell dedifferentiation in *Arabidopsis*. *Curr. Biol.* **21**, 508–514 (2011).
12. Pitaksaringkarn, W. et al. *XTH20* and *XTH19* regulated by ANAC071 under auxin flow are involved in cell proliferation in incised *Arabidopsis* inflorescence stems. *Plant J.* **80**, 604–614 (2014).
13. Asahina, M. et al. Spatially selective hormonal control of RAP2.6L and ANAC071 transcription factors involved in tissue reunion in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **108**, 16128–16132 (2011).
14. Matsuoka, K. et al. Wound-inducible ANAC071 and ANAC096 transcription factors promote cambial cell formation in incised *Arabidopsis* flowering stems. *Commun. Biol.* **4**, 369 (2021).
15. Wu, Y. et al. Rice transcription factor OsDOF11 modulates sugar transport by promoting expression of *Sucrose transporter* and *SWEET* genes. *Mol. Plant* **11**, 833–845 (2018).
16. Zhong, R., Lee, C., Hahighat, M. & Ye, Z.-H. Xylem vessel-specific SND5 and its homologs regulate secondary wall biosynthesis through activating secondary wall NAC binding elements. *New Phytol.* **4**, 1496–1509 (2021).
17. Růžička, K., Ursache, R., Hejátko, J. & Helariutta, Y. Xylem development—from the cradle to the grave. *New Phytol.* **207**, 519–535 (2015).
18. Zhao, B. et al. Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. *BMC*

19. Chan, P. L. et al. Early nodulin 93 protein gene: essential for induction of somatic embryogenesis in oil palm. *Plant Cell Rep.* **39**, 1395–1413 (2020).
20. Schulze, S., Schäfer, B. N., Parizotto, E. A., Voinnet, O. & Theres, K. *LOST MERISTEMS* genes regulate cell differentiation of central zone descendants in *Arabidopsis* shoot meristems. *Plant J.* **64**, 668–678 (2010).
21. Wang, B., Sang, Y., Song, J., Gao, X. Q. & Zhang, X. Expression of a rice *OsARGOS* gene in *Arabidopsis* promotes cell division and expansion and increases organ size. *J. Genet. Genomics* **36**, 31–40 (2009).
22. Lee, S. C., Kim, S. J., Han, S. K., An, G. & Kim, S. R. A gibberellin-stimulated transcript, *OsGASRI*, controls seedling growth and α -amylase expression in rice. *J. Plant Physiol.* **214**, 116–122 (2017).
23. Ueguchi-Tanaka, M. et al. *GIBBERELLIN INSENSITIVE DWARF1* encodes a soluble receptor for gibberellin. *Nature* **437**, 693–698 (2005).
24. Eguchi, S. & Tamura, M. N. Evolutionary timescale of monocots determined by the fossilized birth-death model using a large number of fossil records. *Evolution* **70**, 1136–1144 (2016).
25. Melnyk, C. W. Plant grafting: insights into tissue regeneration. *Regeneration* **4**, 3–14 (2017).
26. Umehara, M. et al. Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **455**, 195–200 (2008).
27. Mylona, P. et al. *Sad3* and *Sad4* are required for saponin biosynthesis and root development in oat. *Plant Cell* **20**, 201–212 (2008).

28. Scarpella, E. & Meijer, A. H. Pattern formation in the vascular system of monocot and dicot plant species. *New Phytol.* **164**, 209–242 (2004).
29. Jura-Morawiec, J., Tulik, M. & Iqbal, M. Lateral meristems responsible for secondary growth of the monocotyledons: a survey of the state of the art. *Bot. Rev.* **81**, 150–161 (2015).
30. Steffens, B. & Rasmussen, A. The physiology of adventitious roots. *Plant Physiol.* **170**, 603–617 (2016).
31. Tillich, H. J. in *Monocots: Systematics and Evolution* (eds Wilson, K. L. & Morrison, D.A.) 603–228 (CSIRO Publishing, 2000).
32. Burger, W. C. The question of cotyledon homology in angiosperms. *Bot. Rev.* **64**, 356–371 (1998).
33. Strosse, H., Houwe, I. & Panis, B. in *Banana Improvement: Cellular, Molecular Biology, and Induced Mutations* (eds Mohan Jain, S. & Swennen, R.) 1–12 (Science Publishers, 2004).
34. Melnyk, C. W. in *Xylem: Methods and Protocols* (eds de Lucas, M. & Etchhells, J. P.) 91–102 (Humana Press, 2017).
35. Hollins, T. W., Scott, P. R. & Gregory, R. S. The relative resistance of wheat, rye and triticale to take-all caused by *Gaeumannomyces graminis*. *Plant Pathol.* **35**, 93–100 (1986).
36. Osbourn, A. E., Clarke, B. R., Lunness, P., Scott, P. R. & Daniels, M. J. An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis* var. *tritici*. *Physiol. Mol. Plant Pathol.* **45**, 457–467 (1994).
37. Chng, S., Crome, M. G. & Butler, R. C. Evaluation of the susceptibility of various grass species to *Gaeumannomyces graminis* var. *tritici*. *N. Z. Plant Prot.* **58**, 261–267 (2005).
38. Löytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis.

Science **320**, 1632–1635 (2008).

39. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
40. Brown, J. W., Walker, J. F. & Smith, S. A. Phyx: phylogenetic tools for unix. *Bioinformatics* **33**, 1886–1888 (2017).
41. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
42. Smith, S. A. & O’Meara, B. C. TreePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
43. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
44. Iwata, N. & Omura, T. Linkage analysis by reciprocal translocation method in rice plants (*Oryza sativa L.*) I. Linkage groups corresponding to the chromosome 1, 2, 3 and 4. *Jpn. J. Breed.* **21**, 19–28 (1971).

Acknowledgements

We thank A. Greenland and E. Wallington for hosting G.R. at the National Institute of Agricultural Botany; F. Gallo and K. Müller for generating electron microscopy images; U. Paszkowski for providing *ccd8* mutant rice; S. Holdgate for providing a culture of *G. graminis* var. *tritici*; M. Castle for statistical advice; M. Tester for supplying tetraploid wheat; R. Bates, R. Donald and K. Billakurthi for assistance; M. J. Grangé-Guermente, O. Murshudova and N. Elina for translating foreign language references into English; and D. C. Baulcombe and A. M. Jones for providing feedback on

this manuscript. G.R. was supported by a Gates Cambridge Trust PhD Student Fellowship; G.R. and P.S. were supported by European Research Council (ERC) Grant 694733 Revolution and BB/P003117/1 awarded to J.M.H.; G.R., A.T. and M.R.W.J. were supported by a Ceres Agri-Tech Fund award; A.K.N. and C.W.M. were supported by a Wallenberg Academy Fellowship (KAW 2016.0274); and C.M. and C.W.M. were supported by an ERC Starting Grant (GRASP-805094).

Author information

Author notes

1. Alison R. Bentley

Present address: International Wheat and Maize Improvement Center (CIMMYT), El Batán, Mexico

Affiliations

1. Department of Plant Sciences, University of Cambridge, Cambridge, UK

Gregory Reeves, Anoop Tripathi, Pallavi Singh, Maximillian R. W. Jones & Julian M. Hibberd

2. NIAB, Cambridge, UK

Gregory Reeves, Melanie Craze, Sarah Bowden & Alison R. Bentley

3. Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

Amrit K. Nanda, Constance Musseau & Charles W. Melnyk

4. Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL, USA

Joseph F. Walker

Contributions

G.R. and J.M.H. conceptualized the study. G.R. A.T., P.S., M.R.W.J., A.K.N. and C.M. performed grafting experiments. A.T. and P.S. prepared sequencing samples and uploaded data to NCBI. G.R., A.T., M.R.W.J., A.K.N. and C.M. performed vascular connectivity assays. G.R. analysed data. M.C. and S.B. generated the transgenic GUS-reporter wheat line. M.C., S.B., A.R.B., P.S., C.W.M. and J.M.H. provided advice on experiments. J.F.W. generated phylogenetic trees. The study was supervised by A.R.B., C.W.M. and J.M.H. The manuscript was written by G.R. and J.M.H. with contributions from the other authors. All authors discussed the results, reviewed and approved the final manuscript.

Corresponding author

Correspondence to [Julian M. Hibberd](#).

Ethics declarations

Competing interests

Cambridge Enterprise has filed a patent with international application no. PCT/GB2019/053232 and publication no. WO/2020/099879, titled ‘Perennial monocotyledon grafting’ and published on 22 May 2020, which includes methods for grafting monocotyledons described in this manuscript. G.R. and J.M.H. are co-inventors on this patent. The other authors declare no competing interests.

Peer review information

Nature thanks Makoto Matsuoka, Colin Turnbull and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Discovery of a method that allows grafting of monocotyledons.

Immature embryos of wheat (*T. aestivum* - orange) and pearl millet (*P. glaucum* - blue) were grown in tissue culture to regenerate a fused plant which may simulate grafting. **a-d**, Grafted plants were not generated after simulating grafting by placing halved calli in contact (**a**), pressing two scutella together in media (**b**), slicing scutellar tissue and placing into close contact (**c**), or placing callus into a toroidal arrangement (**d**). However, after removal and exchange of the central part of immature wheat and pearl millet embryos, some germinated into what appeared to be vestigial grafted plants (**e**). Photographs of tissue (left) are next to graphical representations (right). White arrowheads indicate areas of fusion. Scale-bar represents 0.5 cm and applies to all photographs.

Extended Data Fig. 2 Intra-specific wheat grafts form between different genotypes.

a, Non-grafted wheat seven days after germination. **b**, Section through the root-shoot interface with the vasculature surrounded by green cells. **c**, Self-grafted wheat seven days after plumule transplantation. **d**, Section though graft junction, with the union interrupting the file of green cells linking shoot and root. Parenchyma and scutellar tissues all were connected. **e-h**, Seven day old β -glucuronidase (GUS) wheat (var. Fielder) grafted to wild type (WT) (var. Paragon): grafted seedling (**e**), section through the intra-specific graft junction (**f**), seedling stained with x-gluc (**g**), and ultra-thin section through the stained graft junction (**h**). For **c-h**, *n* is quantified in **k**. **i**, The graft site of a GUS-WT wheat grafted plant four months after grafting and after setting seed. **j**, X-gluc stained section through the junction shows continuous vascular strands connecting scion and rootstock between dashed lines. For **i, j**, *n*=5. **k**, Rates of fusion between intra-specific GUS-

WT grafts one week after grafting, determined destructively by pulling and sectioning (see Supplementary Video 1 for demonstration). Data are presented as box-and-whisker plots displaying median, interquartile range (boxes) and minima and maxima (whiskers) of pooled data from eleven replicate experiments. Comparisons made by two-tailed Fisher's exact test. **I–s**, Vascular connectivity in intra-specific wheat grafts. **I, p, n, r**, The shoot or roots of non-grafted wheat seedlings were inoculated with CFDA-containing agar (**I, p**) or a mock agar (**n** and **r**) solutions seven days after germination as controls to observe vascular transport of fluorescence shoot-to-root (**I**) or root-to-shoot (**p**) 1 h following application. **m, q, o, s**, Even after just 1 h following application, CFDA solutions applied to the shoot or roots of intra-specific GUS-WT wheat grafts identified unsuccessful (**o, s**) or successful shoot-to-root (**m**) and root-to-shoot (**q**) vascular connectivity as shown by empty green arrowheads. GUS-WT grafts successfully transporting CFDA across the graft junction were stained with x-gluc and are shown to the right of the CFDA channel (**m, q**). **t–v**, Root cross-sections of the CFDA phloem connection assay seven days after grafting on intra-specific GUS-WT wheat grafts from **I, m**, and **n**, respectively. **w–y**, Shoot cross-sections of the CFDA phloem connection assay seven days after grafting. Propidium iodide (PI) was applied on the cross-sections (**t–y**), and images were acquired by fluorescence microscopy (**I–y**). Negative controls (**n, r, t, w**) showed very little autofluorescence. For **I–y**, $n=10$ plants evaluated for CFDA movement in grafts and non-graft controls. White arrowheads indicate the graft junction (**c–j, m, q, o, s**). BF, bright field channel. Scale bars represent 1 mm (**a, c, e, g, i**), 500 μm (**I–s**), or 100 μm (**b, d, f, h, j, t–y**).

Source data

Extended Data Fig. 3 Grafted rice forms functional vascular connections.

a–d, Bright-field (BF) images of toluidine blue (TB) or sodium hydroxide (OH^-) cleared mesocotyl tissue in non-grafted (**a, b**) or grafted (**c, d**) rice seven days after grafting. The vascular cylinder is shown between dashed lines. For **a, b**, $n=20$, and for **c, d**, $n=65$. **e–l**, Representative BF and confocal microscopy fluorescence images for exogenous CFDA dye

application indicate movement across graft junctions from shoot-to-root and root-to-shoot in the vasculature. Non-grafted plants transport CFDA from shoot-to-root (**e**) and from root-to-shoot (**i**). Mock solutions generated no CFDA fluorescence signal (**g**, **k**). Grafted plants transport CFDA from shoot-to-root (**f**) and from root-to-shoot (**j**). Lack of CFDA transport is indicative of failed or delayed graft formation (**h**, **l**). **m–r**, Cross-sections indicating CFDA transport to the rootstock (plants from **e**, **f**, and **h**) or scion (plants from **i**, **j**, and **l**) after application of CFDA to the scion or rootstock, respectively, seven days after grafting (grafted five days after germination). All sections were stained with propidium iodide (PI) to visualize shoot or root structure. Images from the CFDA, PI and merged channels are presented. No CFDA fluorescence was detectable in shoots after application of mock agar (**m**). However, when CFDA was provided to roots of non-grafted (**n**) and grafted plants (**o**), signal was detected surrounding xylem tissues of the shoot. Merged images show that CFDA is localized to vascular strands. Enlarged panels show xylem vessels. No CFDA fluorescence was detectable in roots after application of mock agar (**p**), however, when CFDA was provided to shoots of non-grafted (**q**) and grafted plants (**r**), signal was detected in vascular tissue of the root. Merged images show that CFDA is localized to vascular strands. All images acquired by confocal microscopy. **s**, Attachment rate of rice grafts over time (plants from **e–r**). $n=14, 15, 10, 20, 20, 40, 20, 14, 20$ (non-grafted control), $n=30, 30, 60, 50, 59, 240, 120, 40, 80$ (grafted) (1, 2, 3, 4, 5, 6, 7, 10, 14 days after grafting, respectively). **t**, CFDA transport rates in attached grafts from **s** show vascular connectivity increases over the course of seven days. Data are presented as mean grafting rate \pm s.d. of three to six replicate experiments (blocking/random effect). $n=29, 30, 38, 18, 14$. (phloem, non-grafted control) $n=26, 28, 29, 18, 13$ (xylem, non-grafted control), $n=8, 14, 45, 20, 12$ (phloem, grafted), $n=9, 11, 31, 30, 11$ (xylem, grafted) (4,5,6,7,10 days after grafting, respectively). Comparisons were made using mixed effect binomial regression with replicate experiments as a random effect (**s**, **t**). White arrowheads indicate graft junctions (**c**, **d**, **f**, **h**, **j**, and **l**). Scale bars represent 1 mm (**e–l**), 100 μ m (**a–d**), or 50 μ m (**m–r**).

[Source data](#)

Extended Data Fig. 4 Overview of changes in transcript abundance during the development of graft unions in rice.

a, Diagram of non-grafted, wounded, and grafted rice (var. Kitaake) tissue harvested for transcriptome analysis one, three, five, and seven days after grafting. **b**, Three dimensional principle component analysis (PCA) of individual biological replicates for non-grafted, wounded, and grafted rice. $n=3$. **c**, Upset plot for differentially expressed genes (DEGs) for grafted versus non-grafted, and grafted versus wounded rice. **d**, The fifteen most up-regulated GO terms in grafted compared with wounded rice. See Supplementary Table 4 for complete GO term information. **e**, Transcriptional dynamics of genes associated with graft formation. Data are presented as mean \pm s.e.m, and as wounded gene expression overlaps significantly with grafted gene expression during dicotyledon grafting¹⁰, comparisons were made between grafted rice versus non-grafted rice expression by two-way ANOVA. $n=3$. Apart from cambium-maker genes, there was significant overlap of gene expression in rice compared to those associated with dicotyledon grafting.

[Source data](#)

Extended Data Fig. 5 Dynamics of transcripts from cell expansion and cell division genes during graft formation in rice.

a, The fifteen most up-regulated GO terms relative to each adjacent timepoint. Full list of GO details in Supplementary Table 5. **b**, Z-scores for differentially expressed transcripts of cell expansion genes in the small auxin up regulated (SAUR), EXPANSIN and GLYCOSYLHYDROSE families (left), as well as for cell cycle genes (right) in grafted rice over time. Gene details in Supplementary Table 6.

[Source data](#)

Extended Data Fig. 6 Specific gene family analysis for rice graft transcriptomics.

a, Z-score normalized transcription factors with differential transcript abundance within the *APETALA2/ETHYLENE-RESPONSIVE FACTOR (AP2/ERF)*; *NO APICAL MERISTEM, ARABIDOPSIS TRANSCRIPTION ACTIVATION FACTOR, CUP SHAPED COTYLEDON (NAC)*; *MYELOBLASTOSIS (MYB)*; *DNA-BINDING ONE FINGER (DOF)*; and *AUXIN RESPONSE FACTOR (ARF)* families between non-grafted and grafted rice over time as determined by significant Two-way ANOVA grouping variable (graft type, $P < 0.05$). Gene details in Supplementary Table 7. **b**, Pre-microRNA transcripts showing reduced expression during grafting and their potential targets involved in plant development. **c**, Early nodulin genes expression during grafting in rice (above) and *Arabidopsis* during grafting. *Arabidopsis* data was adapted from Melnyk et al. 2018¹⁰. Data are presented as mean \pm s.e.m, and comparisons made by two-way ANOVA (grafted rice versus non-grafted expression) (**b, c**). $n=3$.

[Source data](#)

[Extended Data Fig. 7 Grafting across all three monocotyledonous groups.](#)

a, b, Representative images of non-grafted seedlings (**a**), and hand cross sections of plant tissue (**b**) stained with toluidine blue representing species from Alismatid, Lilioid, and Commelinid monocotyledons. **c, d**, Representative images of grafted seedlings seven to sixty days after grafting (**c**), and hand cross sections through the graft junction stained with toluidine blue (**d**). Panels in the upper right corner show higher magnification of the graft site (**c**). Non-grafted controls are the same age as grafted plants. For **a, b**, $n=5$ and for **c, d**, $n=2$ (*Ananas comosus*), $n=8$ (*Costus laevis*), $n=18$ (*Commelina communis*), $n=8$ (*Beaucarnea recurvata*), $n=16$ (*Dioscorea elephantipes*), $n=12$ (*Gloriosa superba*), $n=12$ (*Arisaema tortuosum*), $n=39$ (*Acorus calamus*). White arrowheads indicate the graft junctions (**c, d**). Scale-bars represent 1 mm (**a, c**) or 200 μ m (**b, d**).

[Extended Data Fig. 8 Intra- and inter-specific graft combinations in the Poaceae.](#)

a, Phylogenetic reconstruction of the twelve subfamilies comprising the Poaceae (grass family). Species grafted are shown in bold. **b**, Rates of fusion between different varieties of the same C₃ or C₄ photosynthetic cereal crop species. **c**, Rates of fusion between different C₃ cereal species.

[Source data](#)

Extended Data Fig. 9 Complementation of rice defective in strigolactone biosynthesis by grafting mutants onto wild-type rootstocks.

a, Position of the *carotenoid cleavage dioxygenase 8* (*OsCCD8*, Os01g0746400) gene in the rice strigolactone biosynthesis pathway. Red color indicates the homozygous mutant *ccd8/ccd8* allele (the d10 mutant⁴⁴). **b**, Height of the tallest ligule (upper plot), and length of the longest leaf (lower plot) on each grafted plant measured over time. Statistically significant differences for tallest ligule height and longest leaf length were assessed by two-way repeated measures ANOVA among graft types over time. Letters indicate significant ($P < 0.05$) groupings among graft types at each time point using pairwise *t*-tests with adjusted P values by the Bonferroni multiple testing correction method. ns = not significant. Data are presented as mean \pm s.d. $n=8$. **c**, Shoot phenotypes of initial grafts (upper panel) and their offspring derived from self-pollination (lower panel) after 40 days of growth in soil. **d**, Rates of reversion to a high tillering (mutant) phenotype in the offspring of self-pollinated grafts between *ccd8* mutant (d10) and wild type (*O. sativa* var. Shiokari). Although the mutant phenotype was rescued when grafted to a wild type root system, offspring of the mutant scions from such grafts reverted to the mutant phenotype, as expected. Scale-bars represent 10 cm (**c**).

[Source data](#)

Extended Data Fig. 10 Grafting wheat to oat confers disease tolerance to take-all.

a, A schematic used for screening wheat, oat, and inter-tribal grafts between wheat and oat for take-all disease, caused by the soil borne pathogen *G. graminis* var. *tritici*. **b**, Representative images of self-grafted oat, self-grafted wheat, and inter-tribal grafts between wheat and an oat. Transverse sections of each graft junction are shown in the upper right corner of each panel seven days after grafting. $n=6$. **c**, Representative images of grafted plants seven days after grafting and immediately after inoculation with an agar plug containing *G. graminis* var. *tritici* ($n = 6$). The plug was placed directly on top of the roots to ensure physical contact between the pathogen and rootstock. **d**, Representative images of grafted plants three weeks after inoculation. Panels show high magnification of graft junction (left), side view of the culture containers (middle), and a view looking down into the containers (right). **e**, Non-inoculated control grafts, with side view of culture containers (left) and view looking down into the container (right). White arrowheads indicate the graft junction (**a–e**), and empty black arrowheads indicate disease progression into the scion past the graft junction (**d**). Scale bars represent approximately 1 mm (**b, d** left panels), 250 μm (**b** upper right panels), 1 cm (**c, d** middle and right, and **e**).

Supplementary information

Supplementary Information

This file contains Supplementary Note 1 (Graft compatibility within monocotyledons) and Supplementary Note 2 (Historical reports of monocotyledon grafting), plus additional references.

Reporting Summary

Supplementary Table 1

Spreadsheet containing total numbers and rates for grafting all monocotyledons in this study. Individual sheets are labelled according to the source data used in each figure of this article.

Supplementary Table 2

Spreadsheet of FPKM values for all genes expressed in across the replicates of grafted, non-grafted, or wounded rice at one, three, five and seven days after grafting. Plots for any gene of interest can be displayed. Up to four plots can be displayed side-by-side for visual comparisons.

Supplementary Table 3

Spreadsheet of all differentially expressed genes between grafted and non-grafted rice, or grafted and wounded rice. Each sheet is separated according to each time point and up- or downregulated genes.

Supplementary Table 4

Spreadsheet of all statistically significant Gene Ontology (GO) terms between grafted and non-grafted rice, or grafted and wounded rice. Each sheet is separated according to each time point and up- or downregulated GO terms.

Supplementary Table 5

Spreadsheet of all statistically significant Gene Ontology (GO) terms between grafted rice at adjacent timepoints during graft union formation partly shown in the bar plots of Extended Data Fig. 5a. Each sheet is separated according to each time point and up- or downregulated GO terms.

Supplementary Table 6

Spreadsheet with gene ID information for the transcript abundance of cell expansion, cell elongation, cell cycle and cell growth regulators in grafted rice shown in the heat map plots of Extended Data Fig. 5b.

Supplementary Table 7

Spreadsheet with gene ID information for the transcript abundance of differentially expressed genes within the AP2/ERF, NAC, MYB, DOF and

ARF transcription factor families between grafted and non-grafted rice shown in the heat map plots of Extended Data Fig. 6a.

Supplementary Table 8

Germination and growth protocols for monocotyledonous species grafted in Fig. 3b.

Supplementary Video 1

Attempted mechanical separation and dissection of a grafted wheat seedling. Real-time footage shows two pairs of forceps attempting to pull apart a grafted intra-specific wheat seedling on wet filter paper one week after grafting. No adhesive or physical attachment were used to aid grafting. The junction did not separate when force was applied in opposing directions from the rootstock and scion. The junction was then bisected laterally with a razor blade to show the fused tissue. This procedure was used to evaluate grafts throughout this study. The video footage was captured at 10 \times magnification with an iPhone SE mounted to a Zeiss Stemi 508 stereomicroscope and was edited using Blender (v.2.80).

Supplementary Video 2

Intra-specific cereal grafting. A demonstration of embryonic shoot (plumule) and embryonic root (radicle) microdissection used to graft different individuals within five cereal species. Following this procedure, seeds are germinated into grafted plants. The video footage was captured on a Leica S8 APO stereomicroscope (Milton Keynes, UK) with a GT Vision GXCAM HICHROME-MET High Resolution Camera (Sudbury, UK) at 10 \times magnification, sped up 2.5 times, and was edited in Blender (v.2.80).

Supplementary Video 3

Inter-specific cereal grafting. A demonstration of embryonic shoot (plumule) and embryonic root (radicle) microdissection used to graft different individuals among five cereal species. Following this procedure, seeds are germinated into grafted plants. The video footage was captured on

a Leica S8 APO stereomicroscope (Milton Keynes, UK) with a GT Vision GXCAM HICHROME-MET High Resolution Camera (Sudbury, UK) at 10 \times magnification, sped up 2.5 times, and was edited in Blender (v.2.80).

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[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. 8](#)

[Source Data Extended Data Fig. 9](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Reeves, G., Tripathi, A., Singh, P. *et al.* Monocotyledonous plants graft at the embryonic root–shoot interface. *Nature* **602**, 280–286 (2022).
<https://doi.org/10.1038/s41586-021-04247-y>

- Received: 14 February 2020
- Accepted: 15 November 2021
- Published: 22 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04247-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

- [**A hard graft problem solved for key global food crops**](#)
 - Colin Turnbull
 - Sean Carrington

Nature (2022)

A hard graft problem solved for key global food crops

- Colin Turnbull
- Sean Carrington

News & Views 25 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04247-y>.

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

- Article
- [Published: 22 December 2021](#)

Morphogen gradient scaling by recycling of intracellular Dpp

- [Maria Romanova-Michaelides](#) ¹ na1,
- [Zena Hadjivasiliou](#) ^{1,2} na1,
- [Daniel Aguilar-Hidalgo](#) ^{1,2},
- [Dimitris Basagiannis](#) ¹,
- [Carole Seum](#) ¹,
- [Marine Dubois](#) ¹,
- [Frank Jülicher](#) ORCID: orcid.org/0000-0003-4731-9185 ^{2,3} &
- [Marcos Gonzalez-Gaitan](#) ORCID: orcid.org/0000-0002-6295-7321 ¹

[Nature](#) volume **602**, pages 287–293 (2022)

- 5470 Accesses
- 113 Altmetric
- [Metrics details](#)

Subjects

- [Body patterning](#)
- [Endocytosis](#)

Abstract

Morphogen gradients are fundamental to establish morphological patterns in developing tissues¹. During development, gradients scale to remain

proportional to the size of growing organs^{2,3}. Scaling is a universal gear that adjusts patterns to size in living organisms^{3,4,5,6,7,8}, but its mechanisms remain unclear. Here, focusing on the Decapentaplegic (Dpp) gradient in the *Drosophila* wing disc, we uncover a cell biological basis behind scaling. From small to large discs, scaling of the Dpp gradient is achieved by increasing the contribution of the internalized Dpp molecules to Dpp transport: to expand the gradient, endocytosed molecules are re-exocytosed to spread extracellularly. To regulate the contribution of endocytosed Dpp to the spreading extracellular pool during tissue growth, it is the Dpp binding rates that are progressively modulated by the extracellular factor Pentagone, which drives scaling. Thus, for some morphogens, evolution may act on endocytic trafficking to regulate the range of the gradient and its scaling, which could allow the adaptation of shape and pattern to different sizes of organs in different species.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

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.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Recycling contributes to gradient formation: photoconversion assay, Dpp transport steps and modules of transports.

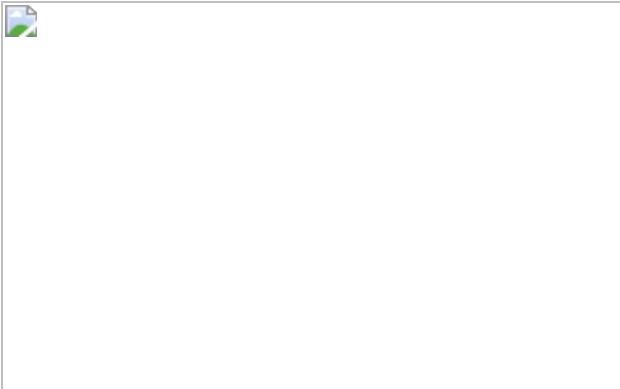


Fig. 2: Parameterization assays.

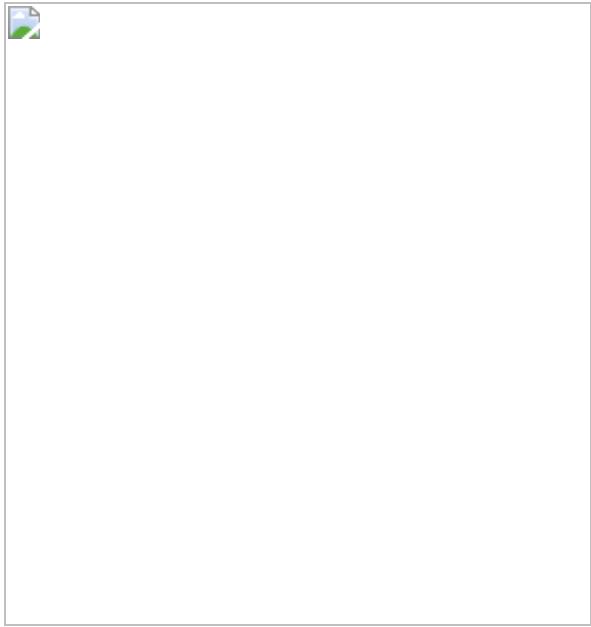


Fig. 3: Parameterization of Dpp transport rates.

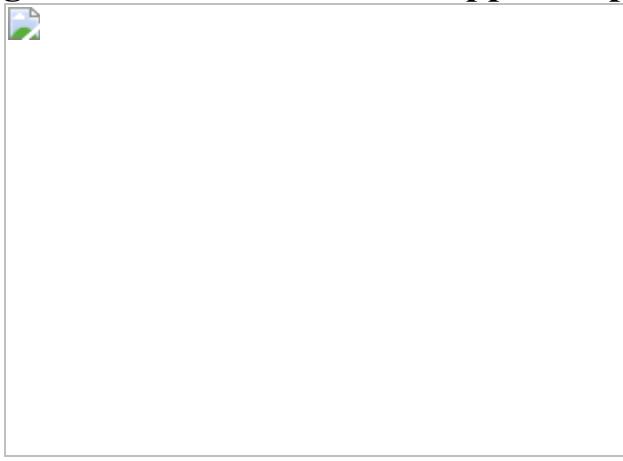
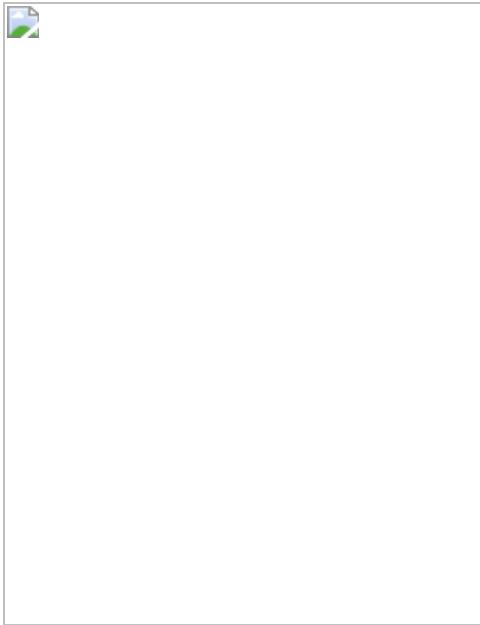


Fig. 4: Internalized Dpp is recycled and spreads.



Data availability

Source data are provided with this paper. Datasets generated during the parameter estimation are available in GitHub (<https://github.com/zenah12/DppTrafficking-/blob/main/README.md>). So
urce data are provided with this paper.

Code availability

Source codes are available in GitHub (<https://github.com/zenah12/DppTrafficking-/blob/main/README.md>). M
ATLAB code corresponding to the binning of control and *pent* mutant data in Fig. 2a is available upon request.

References

1. Rogers, K. W. & Schier, A. F. Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407 (2011).
2. Wartlick, O., Julicher, F. & Gonzalez-Gaitan, M. Growth control by a moving morphogen gradient during *Drosophila* eye development.

Development **141**, 1884–1893 (2014).

3. Wartlick, O. et al. Dynamics of Dpp signaling and proliferation control. *Science* **331**, 1154–1159 (2011).
4. Averbukh, I., Ben-Zvi, D., Mishra, S. & Barkai, N. Scaling morphogen gradients during tissue growth by a cell division rule. *Development* **141**, 2150–2156 (2014).
5. Ben-Zvi, D., Pyrowolakis, G., Barkai, N. & Shilo, B. Z. Expansion-repression mechanism for scaling the Dpp activation gradient in *Drosophila* wing imaginal discs. *Curr. Biol.* **21**, 1391–1396 (2011).
6. Fried, P. & Iber, D. Dynamic scaling of morphogen gradients on growing domains. *Nat. Commun.* **5**, 5077 (2014).
7. Hamaratoglu, F., de Lachapelle, A. M., Pyrowolakis, G., Bergmann, S. & Affolter, M. Dpp signaling activity requires Pentagone to scale with tissue size in the growing *Drosophila* wing imaginal disc. *PLoS Biol.* **9**, e1001182 (2011).
8. Zhu, Y., Qiu, Y., Chen, W., Nie, Q. & Lander, A. D. Scaling a Dpp morphogen gradient through feedback control of receptors and co-receptors. *Dev. Cell* **53**, 724–739 (2020).
9. Inomata, H. Scaling of pattern formations and morphogen gradients. *Dev. Growth Differ.* **59**, 41–51 (2017).
10. Romanova-Michaelides, M., Aguilar-Hidalgo, D., Julicher, F. & Gonzalez-Gaitan, M. The wing and the eye: a parsimonious theory for scaling and growth control? *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 591–608 (2015).
11. Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. Gradient formation of the TGF- β homolog Dpp. *Cell* **103**, 981–991 (2000).
12. Kicheva, A. et al. Kinetics of morphogen gradient formation. *Science* **315**, 521–525 (2007).

13. Stapornwongkul, K. S., de Gennes, M., Cocconi, L., Salbreux, G. & Vincent, J. P. Patterning and growth control in vivo by an engineered GFP gradient. *Science* **370**, 321–327 (2020).
14. Akiyama, T. & Gibson, M. C. Decapentaplegic and growth control in the developing *Drosophila* wing. *Nature* **527**, 375–378 (2015).
15. Matsuda, S. et al. Asymmetric requirement of Dpp/BMP morphogen dispersal in the *Drosophila* wing disc. *Nat. Commun.* **12**, 6435 (2021).
16. Posakony, L. G., Raftery, L. A. & Gelbart, W. M. Wing formation in *Drosophila melanogaster* requires decapentaplegic gene function along the anterior–posterior compartment boundary. *Mech. Dev.* **33**, 69–82 (1990).
17. Aguirre-Tamaral, A. & Guerrero, I. Improving the understanding of cytoneme-mediated morphogen gradients by in silico modeling. *PLoS Comput. Biol.* **17**, e1009245 (2021).
18. Rojas-Rios, P., Guerrero, I. & Gonzalez-Reyes, A. Cytoneme-mediated delivery of hedgehog regulates the expression of bone morphogenetic proteins to maintain germline stem cells in *Drosophila*. *PLoS Biol.* **10**, e1001298 (2012).
19. Roy, S., Huang, H., Liu, S. & Kornberg, T. B. Cytoneme-mediated contact-dependent transport of the *Drosophila* decapentaplegic signaling protein. *Science* **343**, 1244624 (2014).
20. Aguilar Hidalgo, D., Hadjivasilou, Z., Romanova-Michaelides, M., González-Gaitán, M. & Jülicher, F. Dynamic modes of morphogen transport. Preprint at <https://arxiv.org/abs/1909.13280> (2019).
21. Wolpert, M. K. A. L. Mechanisms for positional signalling by morphogen transport: a theoretical study. *J. Theor. Biol.* **191**, 103–114 (1998).
22. Harmansa, S., Hamaratoglu, F., Affolter, M. & Caussinus, E. Dpp spreading is required for medial but not for lateral wing disc growth.

Nature **527**, 317–322 (2015).

23. Felder, S., LaVin, J., Ullrich, A. & Schlessinger, J. Kinetics of binding, endocytosis, and recycling of EGF receptor mutants. *J. Cell Biol.* **117**, 203–212 (1992).
24. Hatta, T. et al. Identification of the ligand-binding site of the BMP type IA receptor for BMP-4. *Biopolymers* **55**, 399–406 (2000).
25. Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb. Perspect. Biol.* **6**, a016725 (2014).
26. Kural, C. et al. Dynamics of intracellular clathrin/AP1- and clathrin/AP3-containing carriers. *Cell Rep.* **2**, 1111–1119 (2012).
27. Mineur, P. et al. Newly identified biologically active and proteolysis-resistant VEGF-A isoform VEGF111 is induced by genotoxic agents. *J. Cell Biol.* **179**, 1261–1273 (2007).
28. Sigismund, S. et al. Clathrin-mediated internalization is essential for sustained EGFR signaling but dispensable for degradation. *Dev. Cell* **15**, 209–219 (2008).
29. Wang, Y., Wang, X., Wohland, T. & Sampath, K. Extracellular interactions and ligand degradation shape the nodal morphogen gradient. *Elife* **5**, e13879 (2016).
30. Waters, C. M., Oberg, K. C., Carpenter, G. & Overholser, K. A. Rate constants for binding, dissociation, and internalization of EGF: effect of receptor occupancy and ligand concentration. *Biochemistry* **29**, 3563–3569 (1990).
31. Allendorph, G. P., Isaacs, M. J., Kawakami, Y., Izpisua Belmonte, J. C. & Choe, S. BMP-3 and BMP-6 structures illuminate the nature of binding specificity with receptors. *Biochemistry* **46**, 12238–12247 (2007).

32. Schwank, G. et al. Formation of the long range Dpp morphogen gradient. *PLoS Biol.* **9**, e1001111 (2011).
33. Hemalatha, A., Prabhakara, C. & Mayor, S. Endocytosis of Wingless via a dynamin-independent pathway is necessary for signaling in *Drosophila* wing discs. *Proc. Natl Acad. Sci. USA* **113**, E6993–E7002 (2016).
34. Zhou, S. et al. Free extracellular diffusion creates the Dpp morphogen gradient of the *Drosophila* wing disc. *Curr. Biol.* **22**, 668–675 (2012).
35. Khmelinskii, A. et al. Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat. Biotechnol.* **30**, 708–714 (2012).
36. Shcherbo, D. et al. Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567–574 (2009).
37. Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nat. Rev. Mol. Cell. Biol.* **2**, 107–117 (2001).
38. Vuilleumier, R. et al. Control of Dpp morphogen signalling by a secreted feedback regulator. *Nat. Cell Biol.* **12**, 611–617 (2010).
39. Fujise, M. et al. Dally regulates Dpp morphogen gradient formation in the *Drosophila* wing. *Development* **130**, 1515–1522 (2003).
40. Akiyama, T. et al. Dally regulates Dpp morphogen gradient formation by stabilizing Dpp on the cell surface. *Dev. Biol.* **313**, 408–419 (2008).
41. Guha, A., Sriram, V., Krishnan, K. S. & Mayor, S. Shibire mutations reveal distinct dynamin-independent and -dependent endocytic pathways in primary cultures of *Drosophila* hemocytes. *J. Cell Sci.* **116**, 3373–3386 (2003).
42. Gupta, G. D. et al. Analysis of endocytic pathways in *Drosophila* cells reveals a conserved role for GBF1 in internalization via GEECs. *PLoS ONE* **4**, e6768 (2009).

43. Mayor, S., Parton, R. G. & Donaldson, J. G. Clathrin-independent pathways of endocytosis. *Cold Spring Harb. Perspect. Biol.* **6**, a016758 (2014).
44. Sabharanjak, S., Sharma, P., Parton, R. G. & Mayor, S. GPI-anchored proteins are delivered to recycling endosomes via a distinct cdc42-regulated, clathrin-independent pinocytic pathway. *Dev. Cell* **2**, 411–423 (2002).
45. Norman, M., Vuilleumier, R., Springhorn, A., Gawlik, J. & Pyrowolakis, G. Pentagone internalises glypicans to fine-tune multiple signalling pathways. *Elife* **5**, e13301 (2016).
46. Yan, D. & Lin, X. Shaping morphogen gradients by proteoglycans. *Cold Spring Harb. Perspect. Biol.* **1**, a002493 (2009).
47. Gratz, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* **196**, 961–971 (2014).
48. Gratz, S. J. et al. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029–1035 (2013).
49. Derivery, E. et al. Polarized endosome dynamics by spindle asymmetry during asymmetric cell division. *Nature* **528**, 280–285 (2015).
50. Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. Hedgehog creates a gradient of DPP activity in *Drosophila* wing imaginal discs. *Mol. Cell* **5**, 59–71 (2000).
51. Eugster, C., Panakova, D., Mahmoud, A. & Eaton, S. Lipoprotein-heparan sulfate interactions in the Hh pathway. *Dev. Cell* **13**, 57–71 (2007).
52. Marois, E., Mahmoud, A. & Eaton, S. The endocytic pathway and formation of the Wingless morphogen gradient. *Development* **133**, 307–317 (2006).

53. Loubry, S. & Gonzalez-Gaitan, M. Monitoring notch/delta endosomal trafficking and signaling in *Drosophila*. *Methods Enzymol.* **534**, 301–321 (2014).
54. Montagne, C. & Gonzalez-Gaitan, M. Sara endosomes and the asymmetric division of intestinal stem cells. *Development* **141**, 2014–2023 (2014).

Acknowledgements

We thank R. Mateus and I. Castanon Ortega for their feedback on the manuscript; F. Karch for Cre recombinase; K. Basler for the stock to make *brk^{M68} tkv⁸* mutant clones; E. Doná for the p744_p3E_mkate2sfGFP plasmid; M. Affolter for the LexA inducible eGFP::DPP; K. Kruse for discussions; E. Derivery for the purified sfGFP-mKate2; V. Rasul-Kareeva for various contributions; and the Bioimaging Center of University of Geneva for microscopy support. Z.H. was supported by an HFSP long-term fellowship; and M.G.-G. was supported by the DIP of the Canton of Geneva, SNSF, the SystemsX EpiPhysX grant, the ERC (Sara and Morphogen) and the NCCR Chemical Biology program.

Author information

Author notes

1. These authors contributed equally: Maria Romanova-Michaelides, Zena Hadjivasiliou

Affiliations

1. Department of Biochemistry, Faculty of Sciences, University of Geneva, Geneva, Switzerland

Maria Romanova-Michaelides, Zena Hadjivasiliou, Daniel Aguilar-Hidalgo, Dimitris Basagiannis, Carole Seum, Marine Dubois & Marcos Gonzalez-Gaitan

2. Max Planck Institute for the Physics of Complex Systems, Dresden,
Germany

Zena Hadjivasiliou, Daniel Aguilar-Hidalgo & Frank Jülicher

3. Cluster of Excellence Physics of Life, TU Dresden, 01307 Dresden,
Germany

Frank Jülicher

Contributions

M.R.-M. performed most experiments and quantifications and performed data analysis. D.A.-H., Z.H. and F.J. developed the theory and performed data analysis. Z.H. and D.A.-H. performed numerical simulations. C.S. cloned and made fly stocks to express Dpp^{Timer} and UAS Pentagone::GFP, performed immunoprecipitations and purified the GBP. D.B. performed the photoconversion experiment, labelled the purified GBP with Alexa555, cloned and purified GBP–Dendra2 and developed the acid wash. M.D. made the Dpp^{CRISPR} stock. M.R.-M. and M.G.-G. conceived and designed experiments. M.R.-M. and Z.H. prepared figures. M.R.-M., Z.H., F.J. and M.G.-G. prepared the manuscript.

Corresponding authors

Correspondence to [Maria Romanova-Michaelides](#), [Frank Jülicher](#) or [Marcos Gonzalez-Gaitan](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Dagmar Iber, Timothy Saunders and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Photoconversion assay controls and two extreme regimes of Dpp transport.](#)

a–d, Photoconversion assay. Test of efficiency of the acid wash in the photoconversion experiment using GBP-Dendra2: GFP-Dpp expressing discs have been incubated in GBP-Dendra2 for 50 min at 4 °C (the nanobody is only bound to the extracellular pool) and subsequently acid-washed to remove the label of the extracellular pool. Confocal image of eGFP-Dpp^{LOP} expressing disc (**a**) and corresponding images of Dendra2* (**b**) before photoconversion (left) and 40 min following photoconversion (right; see Materials and Methods). Note that no detectable Dendra2* signal is observed 40 min after the acid wash, indicating that the extracellular pool of nanobodies has been efficiently removed and that the potential extracellular leftover (below the detection limit) cannot lead to an observable recovery in intracellular compartments. **c**, Comparison between eGFP-Dpp^{Gal4} gradient profiles and gradient profiles formed by photoconverted Dendra2* propagated into the posterior compartment of the discs (photoconversion experiments as in Fig. [1a](#)). Bar plot showing $\phi = \lambda/l$ of eGFP-Dpp^{Gal4} gradient profiles and photoconverted Dendra2* gradient profiles for large discs. Bars, standard deviations. Two-tailed two sample t-test, p-value = 0.2353. **d**, Fluorescence intensity of Dendra2* in a ROI of 6x35 μm at the source boundary in the photoconversion experiment in Fig.

1a. Measured Dendra2* fluorescence (blue dots) is plotted as a function of time after the photoconversion event. The red line represents the theoretical dynamics of Dendra2* fluorescence signal considering the parameterized values for large discs. $n = 4$ biologically independent samples. Data represented as mean values \pm s.e.m. **e–h**, Acid wash efficiently removes the extracellular pool. Confocal images of eGFP-Dpp^{LOP} gradient (green in **e**, **f**), and extracellular eGFP-Dpp^{LOP} pools monitored by means of an extracellular immunostaining (see Materials and Methods, [Supplementary Information section 2.3.2](#)) by using a GBP-Alexa555 nanobody against GFP (**g**, **h**; red in **e**, **f**) before (**e**, **g**) and after (**f**, **h**) acid wash. Acid wash in these conditions largely reduces the extracellular staining down to 9% of the signal. Scale bar: 10 μm . **i**, Acid wash does not affect internalized GBP-Alexa555. Confocal images of eGFP-Dpp^{LOP} (top, green) and GBP-Alexa555 internalized for 40 min (bottom, red) before (left) and after acid wash (right). The GBP-Alexa555 signal decreases by $2.3 \pm 0.6\%$ after acid wash. **j**, Acid wash: effect of pH on GBP binding to GFP from larval extracts. Immunoblot of GFP which was bound to GFP-Trap beads (Chromotek, GFP-Trap beads, lanes 3-7) and GFP dissociated from GFP-Trap beads (supernatant, lanes 8-12) following treatment at different pH. FT, flowthrough (lane 1), PD, pulldown (lane 2). For gel source data, see [Supplementary Fig. 1a](#). **k**, Stacked bar chart showing the relative contribution of the different modules to Dpp transport in the two theoretical extreme regimes of morphogen transport: extracellular diffusion (ExD²⁰) and transcytosis (Tr) regimes. The relative contribution of different modules is expressed as the ratio λ_i^2/λ^2 with the index i corresponding to each of the four modules ($i = u,b,r,t$). Note that the unbound module contributes almost exclusively to λ^2 in ExD and the transcytosis module, in Tr. **l**, Theoretical values of the 8 transport rates characteristic for ExD (rate values as in reference²⁰) and Tr regimes of morphogen transport. **m**, **n**, FRAP recovery with respect to the two extreme theoretical regimes. Red lines, calculated recovery curves in a FRAP experiment for a set of parameter values corresponding to the extreme Tr (**m**) and ExD¹⁷ regimes (**n**). Blue dots, average of the experimental recovery curves in discs of $l = 144 \mu\text{m}$ average posterior length. $n = 9$ biologically independent samples. Data represented as mean values \pm s.e.m. The coefficient of determination R^2 characterizes how well the calculated curves fit the experimental FRAP data. λ , decay

length of the Dpp gradient profile calculated using equation (1) and the set of parameter values corresponding to Tr and ExD (see [Supplementary Information section 4.2](#)). Bars, s.e.m. Scale bar, 10 μm (**a, h, i**).

[Source data](#)

[Extended Data Fig. 2 Analysis of Dpp leakage and effects of growth on Dpp gradient profile.](#)

a, Confocal images of GBP-Alexa555 labelling extracellular GFP-Dpp in a control extracellular staining (top) and following a chase of living discs for 7 h at 4 °C. **b**, Total GBP-Alexa555 fluorescence in the conditions in **a**. Two-tailed two sample t-test, p-value = 0.7787. n, number of biologically independent samples. Bars, s.e.m. **c, d**, Schemes of sGFP^{Ptc-Wg}, see reference³¹ (**c**) and sGFP^{Dpp} constructs (**d**). Sizes of fragments represented in the scheme do not correspond to the nucleotide sequences. **e**, Confocal images of sGFP^{Dpp} (top), phalloidin staining (middle) and merge (bottom). Left panels, orthogonal views; right panels, xy plane. **f**, Normalized average spatial profile of sGFP^{Dpp} fluorescence (green) compared to the normalized profiles of gradients with decay lengths $\lambda=\lambda_{\text{Dpp}}$; $\lambda = 6L$; $\lambda = 3L$ and $\lambda = 2L$ with $\lambda_{\text{Dpp}} = 28.9 \mu\text{m}$ and $L = 144.6$, average posterior size of eGFP-Dpp^{LOP} third instar discs. **g**, Orthogonal views of confocal images of sGFP^{Dpp} fixed immediately after dissection (0 h) and following a chase of living discs for 1h at 25 °C and 4 °C. **h**, Total sGFP fluorescence in the conditions in **g**, normalized for each temperature to the value at $t = 0$ h. Two-tailed two sample t-test for unequal variances, p-values: 0.9792 (25 °C) and 0.7543 (4 °C). n, number of biologically independent samples. Bars, s.e.m. **i**, Effect of leakage on parameterization of Dpp transport rates. Average estimated parameters considering leakage rates $k_L = 0\text{s}^{-1}$; 0.00001s^{-1} ; 0.0005s^{-1} and 0.001s^{-1} . Simulations represent 3.7×10^6 randomly chosen parameter sets per condition. **j**, Stacked bar chart showing the relative contribution of the different modules to λ^2 (described in Fig. [1e,f](#)) for conditions in **i**. n, sample size; bars, s.d. **k**, Long-term FRAP assay. Dynamics of fluorescence recovery in conventional FRAP for one hour (red) and long-term FRAP for ten hours (blue). Fluorescence recovery is normalized to the signal in the ROI before bleaching. Note that recovery of conventional FRAP overlays

the dynamics of long-term FRAP at short time scales. Bars, s.e.m. **l**, **n**, Dynamics of long-term FRAP recovery and fit to double (**l**, blue line) and single exponential dynamics (**n**, blue line) to the dataset (both early and late). Box in **l**, late recovery (after 5,000 s) analysed in **m**. **m**, Dynamics of long-term FRAP recovery (late recovery) and single exponential fit (blue line) to the late slow dynamics. **o**, Wing disc area plotted as a function of disc age in staged larvae (hours after egg laying) and fit to an exponential growth in which growth rate decays exponentially over time (red line). See [Supplementary Information section 2.9](#). Orange and blue lines correspond to area and age of discs of $l = 144 \mu\text{m}$ and $l = 80 \mu\text{m}$ posterior length, respectively, as determined by the plot in **p**. **p**, Posterior compartment length (l) as a function of wing disc area (A). Black line, power-law fit. Growth anisotropy $\langle m = g_x \rangle / g = \frac{\dot{A}}{A} / \frac{\dot{l}}{l}$. Using m , the area of discs of $l = 144 \mu\text{m}$ and $l = 80 \mu\text{m}$ posterior length can be determined (orange and blue lines). **q**, Wing disc growth rate (g), relaxation rate of the slow dynamics (that of the immobile fraction, IF) in long-term FRAP (k_{IF}) and degradation rate of the immobile pool (k_2) estimated according to $k_2 = k_{\text{IF}} - g$. The timescales corresponding to these rates are indicated on top of bars. **r**, **s**, Measurement of the mobile pool decay length. **r**, Confocal images of eGFP-Dpp^{LOP} before (top) and at indicated times after bleaching (middle and bottom). **s**, Correlation between the decay length of the total pool of eGFP-Dpp at steady state (λ_T) measured before bleaching and the mobile pool decay length measured 30 min after bleaching (λ_M). Black line, linear regression. Note the slope close to 1, indicating that for discs of different sizes $\lambda_M \approx \lambda_T$. Scale bar, 10 μm (**a**, **e**, **g**, **r**).

[Source data](#)

[Extended Data Fig. 3 Parameterization assay controls I: steady-state decay length and nanobody internalization.](#)

a, Immunoprecipitation of eGFP-Dpp under different expression systems. See [Methods](#). Input (I) and immunoprecipitate (IP) from eGFP-*Dpp*^{CRISPR/+} (lanes 1,2), eGFP-*Dpp*^{CRISPR/CyO, Dpp⁺ (lanes 3,4), *dppLG/+*; *LOP-eGFP-Dpp/+* (lanes 5,6; eGFP-Dpp^{LOP}) and *Dpp-Gal4/UAS-*}

sfGFP-mKate2-Dpp larval head extracts (lanes 7,8). Mature GFP-Dpp fragment after Furin cleavages is marked by an asterisk. Note that GFP-Dpp amounts when expressed using LexA/LOP system are similar to the amounts of GFP-Dpp endogenously expressed (1.1 fold), whereas Gal4/UAS system expresses almost 400 fold more GFP-Dpp. For gel source data, see [Supplementary Fig. 1b](#). **b**, Confocal image of eGFP-Dpp^{LOP} in the background of overexpression of Dpp by dppGal4. **c, d**, Dynamics of FRAP recovery (**c**) and nanobody uptake (**d**) in this condition (red lines) as compared to control (blue). Bars, s.e.m. **e**, Average decay length λ of the gradients considered in the three datasets, corresponding to the three conditions considered in this report: large discs (average posterior length $l = 144 \mu\text{m}$ in the dataset), small discs (average $l = 80 \mu\text{m}$) and in a *pent*² mutant disc (average $l = 130 \mu\text{m}$). Bars, standard error to the mean (s.e.m.). The average decay length for the average l corresponding to the three experimental conditions was estimated using the linear regression of eGFP-Dpp^{LOP} control (sample size $n = 157$ discs) and *pentagone* mutant ($n = 63$ discs) datasets (see Fig. [2a](#)). **f**, Confocal images (maximum projections) of the eGFP-Dpp^{LOP} gradient (red box, region of interest (ROI) in the posterior compartment) in representative discs from the three conditions described in **b**. The source is to the left. **g**, Average spatial distribution of eGFP-Dpp^{LOP} in these datasets. Shaded areas, s.e.m. Black line, exponential fit. **h, i**, Left, normalized eGFP-Dpp^{LOP} profiles in large control discs (**h**; $l = 144 \mu\text{m}$) and *pent*² mutant disc (**i**; $l = 130 \mu\text{m}$); right, average residuals of the fits of these profiles to an exponential function. Bars, s.e.m. **j**, Scaling plot of eGFP-Dpp^{LOP}. Decay length (λ , from the exponential fit) of the eGFP-Dpp^{LOP} gradient versus l . Red line, linear regression. $\phi_L = \lambda/l$ determined from the linear regression. **k**, GBP-Alexa555 signal intensity as a function of time in 13 different discs. Lines, fits to the phenomenological $\langle c \rangle_{\rm rm}^{\langle T \rangle} \langle i \rangle(t)$ equation for the internalized signal intensity (left equation in **m**; red/green boxes as in **l**). **l**, Average dynamics of the GBP-Alexa555 fluorescence signal in the three conditions. Bars, s.e.m. **m**, Parameterization of k_N , k_o and k_r based on the dynamics of GBP-Alexa555 signal. Left, phenomenological $\langle c \rangle_{\rm rm}^{\langle T \rangle} \langle i \rangle(t)$ equation which captures the exponential (red box; see also **l**) and linear dynamics (green box) of the accumulation of the GBP-Alexa555 signal. Right, relationship between the phenomenological parameters A, B and p and k_N , k_o and k_r

(see [Supplementary Information section 2.2.1](#)). **n**, Scheme of the GBP-Alexa555 internalization assay. Rates and pools indicated, like in Fig. [1d](#). Note that the fluorophore (Alexa555; star) degrades on a time scale which is much longer than the duration of the experiment. **o**, Confocal images of internalized GBP-Alexa555 in a disc expressing eGFP-Dpp^{LOP} (top) and a control disc (bottom) at indicated timepoints of nanobody internalization using the same nanobody concentration as in Fig. [2b–f](#). Note that, under these conditions, fluid-phase internalization of the nanobody in the absence of eGFP-Dpp^{LOP} (bottom, control) is negligible compared to the internalization when bound to eGFP-Dpp^{LOP} (top, eGFP-Dpp). **p**, Dynamics of internalized GBP-Alexa555 in the disc expressing eGFP-Dpp^{LOP} (green curve) and a control disc (blue curve), in the same experimental conditions (e.g. same nanobody concentration) as in the nanobody uptake experiments in **o**. Note that, in these conditions, internalization of GBP-Alexa555 by fluid phase in the absence of GFP-Dpp is negligible. **q–r**, Dynamics of fluid-phase internalization of GBP-Alexa555. **q**, Confocal image of fluid-phase internalized GBP-Alexa555 (40 min of nanobody incubation) showing that, at high concentration of the nanobody, a signal can be detected at low levels which is homogenous in space (there is no gradient). Five-fold higher concentration of the nanobody than in **o** was used to reliably detect the signal of the fluid-phase internalized nanobody. **r**, Dynamics of fluid-phase internalized GBP-Alexa555 signal intensity, averaged over 3 independent experiments. Same concentration as in **p**. Shaded area, s.e.m. Note that the dynamics do not show the early exponential regime seen in the presence of eGFP-Dpp, indicating that the nanobody by itself is not significantly recycled. **s**, Top, confocal image of fluid-phase internalized Alexa555 (40 min of Alexa555 incubation). Also here, internalization of the fluorophore is homogeneous in space. Bottom, high magnification of the ROI area shown in the top. **t**, Dynamics of fluid-phase internalized Alexa555, showing a linear increase without saturation in the timescale of the experiment, which reflects a lack of degradation in the lysosome of the Alexa555 fluorophore. **u**, Confocal images of the eGFP-Dpp^{LOP} gradient (left) and internalized GBP-Alexa555 (right) after 45 min of incubation with the nanobody in a control large disc. The source is to the left. In contrast to the situation for fluid phase internalization (**p**, **r**), internalized eGFP-Dpp^{LOP} with GBP-Alexa555 is distributed as a gradient. **v**, Spatial profiles of the gradients in **u** in the

posterior compartment. The decay length is determined by fitting the spatial profiles to an exponential function with an offset. The decay length is given with its confidence interval. n, number of biologically independent samples. Bars, s.e.m (**c**, **g**, **h**, **l**, **r**). Scale bars, 10 μm (**b**, **f**, **o**, **s**, **u**) and 50 μm , (**q**).

Source data

Extended Data Fig. 4 Parameterization assay controls II: FRAP, extracellular fraction determination and parameter estimation by ABC.

a, Left, confocal image of the eGFP-Dpp^{LOP} gradient in a FRAP experiment (source and posterior compartment). Red box, region to be photobleached. Right, eGFP-Dpp^{LOP} fluorescent signal in the red box region before photobleaching (-1 min) and at different times (as indicated below) after photo-bleaching. **b**, Average dynamics of fluorescence recovery in the bleached area in the three experimental conditions (discs of $l = 144 \mu\text{m}$ and $l = 80 \mu\text{m}$ posterior length and in a *pent*² mutant disc). Data represented as mean values. Bars, s.e.m. Lines, calculated recovery using the five-pool theoretical framework for a set of parameter values. The coefficient of determination R^2 characterizes how well the theoretical curves fit the FRAP data. n, sample size. **c**, **d**, Robustness analysis of the FRAP assay. The average FRAP trace was fitted by a single dynamic equation³. Dependence of the goodness of the fit (R^2) to this single dynamic equation (**c**) and the effective diffusivity (D_{eff}) estimated by this fit (**d**) on the number of individual recovery curves (n) considered for the average FRAP trace. The analysis was performed for the three experimental conditions of this report: large discs (average posterior length $l = 144 \mu\text{m}$ in the dataset; left), small discs (average $l = 80 \mu\text{m}$; centre) and *pent*² discs (right). Bars, confidence intervals (**d**). In **d** data are represented as D_{eff} estimated by fit for varying number of independent recovery curves, n. Bars, confidence intervals of fit. **e**, Effective diffusivity (D_{eff} , left) and effective degradation rate (k_{eff} , right) plotted against the average posterior length of discs within two datasets: small (average $l = 80 \mu\text{m}$) and large (average $l = 144 \mu\text{m}$). The average FRAP recovery curve was fitted by a

single dynamic equation³ to determine D_{eff} and k_{eff} . Note, that as discs grow, D_{eff} does not change significantly, whereas k_{eff} decreases significantly, as previously reported²³. Data is represented as D_{eff} and k_{eff} estimated by fit. Bars, confidence intervals of fit. n, number of biologically independent samples. One-tailed two sample t-test with unequal variances; p-values: 0.1765 (D_{eff} , left) and 0.0038 (k_{eff} , right). **f**, Simulated intensity profile of eGFP-Dpp^{LOP} at indicated times after photobleaching in the ROI in the posterior compartment (experiment as in **a**). x, distance from the edge of the anterior compartment. Parameter values used in the simulations are those of our parameterization for $l = 144 \mu\text{m}$. **g**, Confocal images of the eGFP-Dpp^{LOP} gradient (left; total pool), and the extracellular eGFP-Dpp^{LOP} pools monitored by means of an extracellular immunostaining (see [Supplementary Information section 2.3](#)) by using a GBP-Alexa555 nanobody against GFP (right; extracellular pool). Higher magnification of the fluorescent signal of the area boxed in the images are shown to the right. **h**, Expression of the extracellular fraction (ρ) as function of Dpp transport rates. **i**, Equimolarity of the GBP-Alexa555 and eGFP solutions used for calibration of the Alexa555 *versus* GFP fluorescent signal (see [Methods](#), [Supplementary Information section 2.3.2](#); relevant to the extracellular fraction determination assay). The concentrations of GBP-Alexa555 and eGFP was first roughly determined by means of a BCA assay ([Supplementary Information section 2.3.2](#)). Plot of GFP fluorescence intensity as a function of the ratio of GBP-Alexa555 and GFP concentrations (determined by BCA) in the solutions. The relative concentration of GFP and GBP-Alexa555 can be determined from the relative concentration at which the minimum value (r_{\min}) of GFP fluorescence has been reached. Note that $r_{\min} \approx 1$ confirms that the BCA estimation was already accurate. **j**, Parameter value sets determined by the parameterization procedure (see [Supplementary Information section 2.5.2](#)) are represented in the ($k_{\text{on}}, k_{\text{off}}$) plane. Light orange area represents the full space of 3×10^7 parameter value sets considered ($l = 144 \mu\text{m}$ dataset). Dark orange dots represent sets of parameter values within those which satisfy the constraints given by the steady-state decay length, the long-term FRAP assay, the nanobody internalization and the FRAP assay. Calculated FRAP recovery curves using these sets of values fit the experimental FRAP data

with $R^2 > 0.92$. Note that the solutions are separated into two clusters (clouds): the upper cloud, with higher k_{on} , k_{off} , is characterized by a low extracellular fraction $\rho < 0.10$ and a lower cloud, by a high $\rho < 0.25$. **k**, Selected sets of parameter values from **j** for which the calculated extracellular fraction is within the experimentally determined range of ρ values ($0.08 < \rho < 0.18$). **I**, Sets of parameter values which satisfy all the constraints given by our assays (see [Supplementary Information section 2.5.2](#)), represented in (k_{off}, k_{on}) , (k_{off}, k) , (D_0, k_{on}) and (k_o, k_{on}) planes. The parameter values corresponding to the two extreme theoretical cases discussed in [Supplementary Information section 4.2](#) (Extracellular diffusion regime, ExD, yellow and Transcytosis regime, Tr, purple) are represented by circles for comparison. **m**, Average estimated parameters in the three experimental conditions compared to the theoretical values of parameters in ExD and Tr. Bars, s.d. N, number of parameterized sets of values. Scale bars: 10 μm (**a**, **g**).

[Source data](#)

[Extended Data Fig. 5 Quantitative considerations: robustness analysis and decay length boosts.](#)

a, Cluster of parameter value sets in the (k_{on}, k_{off}) plane corresponding to three different ranges of R^2 to the experimental FRAP recovery for the three experimental conditions. The coefficient of determination R^2 characterizes the goodness of the fit between the FRAP data and the calculated recovery curves. Relaxing the quality of fit down to $R^2 > 0.85$ (from $R^2 > 0.93$) does not populate the lower cloud, and therefore does not affect the assignment to the ExD-type versus Combined transport regimes. Points that populate the lower cloud as in the $l = 80 \mu\text{m}$ and $pent^2$ conditions) require that $R^2 < \sqrt{R_{th}^2 - 2}$ (see [Supplementary Information section 3.7](#) for details). **b**, Cluster of parameter value sets in the (k_{on}, k_{off}) plane corresponding to different ranges of calculated extracellular fraction ρ for the three experimental conditions. An increase in ρ beyond ρ^* is required to shift the solutions to the “lower” cloud. The lower cloud is characteristic of the ExD-type regime. **c**, **d**, Sets of parameter values (clouds of points) compatible with all the assays considered in this report in the (k_{on}, k_{off})

plane. Isolines for Boost k_r (**c**) and Boost k_{off} (**d**) are also represented (see look up table). See [Supplementary Information section 3.5](#) for definition of the Boosts. The three conditions considered in this work are shown: large discs (average posterior length $l = 144 \mu\text{m}$ in the dataset; left), small discs (average $l = 80 \mu\text{m}$; centre) and *pent*² discs (right). **e**, Average calculated Boost k_r , Boost k_{off} and Boost D_0 for the three experimental conditions compared to the calculated Boosts for the theoretical values of parameters in the ExD and Tr regimes. N, number of parameterized sets of values. Data represented as mean values over N parameterized value sets. Bars, s.e.m. **f–i**, iFRAP assay. **f**, Scheme of the iFRAP assay (see [Supplementary Information section 2.7](#)). **g, h**, Test of efficiency of the acid wash in the iFRAP (and photoconversion) experiment: GFP-Dpp expressing discs have been incubated in GBP-Alexa555 for 50 min at 4 °C (the nanobody is only bound to the extracellular pool) and subsequently acid-washed to remove the label of the extracellular pool. Confocal image of eGFP-Dpp^{LOP} expressing disc (**g**) and corresponding images of GBP-Alexa555 (**h**) at indicated times after the acid wash (see Materials and Methods). Note that no detectable GBP-Alexa555 signal is observed 40 min after the acid wash, indicating that the extracellular pool of nanobodies has been efficiently removed and that the potential extracellular leftover (below the detection limit) cannot lead to an observable recovery in intracellular compartments. **i**, Theoretical dynamics of GBP-Alexa555 fluorescence recovery in the iFRAP experiment normalized to the pre-photobleaching levels. Recovery was calculated numerically using the set of values determined experimentally for large (top) and small discs (bottom). The dashed lines indicate the estimated fraction of recovery 2,000s after photobleaching in large and small discs to compare with the experimental conditions in the iFRAP experiments (Fig. 4g). Scale bar, 10 μm (**g, h**).

[Source data](#)

[Extended Data Fig. 6 Internalized Dpp is recycled and spreads in the tissue: Dpp^{Timer} and recycling Rab proteins.](#)

a, Functionality of Dpp^{Timer}. Left, control disc, expressing sfGFP-mKate2-Dpp under the control of the GAL4/UAS expression system (Dpp^{Timer}).

Centre, *dpp* mutant disc, the wing imaginal disc is outlined with the white dashed line. Right, *dpp* mutant disc expressing Dpp^{Timer}. Note that the mutant phenotype seen in the central image is rescued. **b**, Scatter plot of sfGFP and mKate2 pixel intensities and linear fit to obtain the calibration factor F (see [Supplementary Information section 2.6.3](#)). n = 23 beads. **c**, Confocal images of the Dpp^{Timer} gradient in the wing disc (sfGFP, top and mKate2, bottom). **d**, Relative concentration profiles of mature sfGFP and mKate2 plotted against the distance from the Dpp source (see [Supplementary Information section 2.6.3](#)), corresponding to the intensity profiles measured from the images in **c**. These intensity profiles represent the relative amounts of sfGFP and mature mKate2 molecules. **e**, Adjusted fluorescence intensity profiles for sfGFP ($g^*(x)$) and mature mKate2 ($r^*(x)$) which are proportional to the respective concentration profiles. X-axis represents the distance from the source. Red dashed line is positioned at the anterior-posterior boundary. Note that both in the source and in the region of the target closer to the source, there are less mature mKate2 molecules, confirming that Dpp molecules are younger closer to the source. **f**, Plotted relative age ($A(x)$) of Dpp molecules as a function of position calculated from the calibrated profiles in **e**. Note that as molecules move away from the source they become older on average: $A(x)$ increases to plateau at values close to 1. n, number of biologically independent samples. Shaded areas, s.e.m (**e**, **f**). **g–j**, Effect of pH on the Timer. **g**, Control of the bafilomycin treatment. Confocal images of a ROI in discs incubated with a LysoSensor™ probe for 30 min before (top) and after (bottom) incubation in control Clone 8 medium (right) or bafilomycin solution (left). **h**, Effect of pH on sfGFP and mKate2 in the Dpp^{Timer}. Confocal images of sfGFP (left) and mKate2 (right) of Dpp^{Timer} before (top) and after (bottom) neutralization of pH to 7 following bafilomycin treatment for 30 min. **i**, Fluorescence signal decrease of sfGFP and mKate2 owing to acidic pH in intracellular compartments. Percentage decrease of fluorescence from pH 7 (discs after bafilomycin treatment) to the acidic environment in intracellular compartments (discs before bafilomycin treatment). Note that the decrease is very similar for both fluorophores. **j**, Normalized fluorescence intensity of sfGFP (blue) and mKate2 (orange) in purified Timer molecules in solutions at different pH. Data normalized to the intensity at pH 7.4. The number of biologically independent samples for

this analysis: $n_{\text{pH}5.86} = 8$; $n_{\text{pH}6.4} = 7$; $n_{\text{pH}7.4} = 7$; $n_{\text{pH}7.9} = 5$. Data represented as mean values \pm s.e.m. Note, that the difference between the normalized intensity of sfGFP and mKate2 at the different pH value is not significant ($p\text{-value}>0.05$; two-tailed two sample t-test). **k**, Confocal images of eGFP-Dpp^{LOP} in control condition (top) and after RNAi through expression of dsRNA for the recycling Rab proteins, Rab11 (middle) and Rab4 (bottom) in posterior target cells. **l**, Spatial fluorescence profiles of eGFP-Dpp^{LOP} corresponding to control (top), Rab11RNAi (middle) and Rab4RNAi (bottom) conditions in **k**. **m**, Decay length λ of eGFP-Dpp^{LOP} gradient versus posterior compartment length l for control ($n = 157$), *pent*² discs ($n = 63$) and Rab4RNAi ($n = 39$). Dots, binned data; bars, s.e.m. Control and *pent*² data as in Fig. 2a, Extended Data Fig. 7f. **n**, Average eGFP-Dpp^{LOP} decay length in control and Rab11RNAi conditions. Difference between the two conditions is significant as determined by a two-tailed, two sample t-test with unequal variances, $p\text{-value} = 0.0034$. **o**, Recycling rate in control and Rab4RNAi conditions, determined by the nanobody uptake assay. Number of curves for each condition is $n = 4$. Difference between the two conditions is significant; two-tailed, two sample t-test with unequal variances, $p\text{-value} < 0.0001$. Rab4RNAi expression was driven by means of the thermosensitive Gal4Gal80^{ts} system (29 °C). **p–r**, Scaling of eGFP-Dpp^{LOP}. **p**, Dpp gradient profiles of discs from 40 to 160 μm posterior length. Each individual profile was fitted to an exponential function with an offset (see [Supplementary Information section 2.1.2](#)) and the offset returned from the fit was subtracted. **q**, Normalized Dpp gradient profiles. Each profile was normalized to the amplitude C_0 of its exponential fit in the ordinates ($C(r)/C_0$) and to the posterior length l of the corresponding wing disc in the abscissas ($r=x/l$). Shaded area, s.e.m. Black line, average normalized profile. **r**, Density plot of **q**: Colour-code corresponds to the fraction of the number of gradients passing through a certain r , $C(r)/C_0$ bin. Scale bars, 100 μm (**a**) and 10 μm (**c, g, h, k**).

[Source data](#)

[Extended Data Fig. 7 Gradient scaling by recycling: Pentagone.](#)

a, Continuous and monotonic transition from $\lambda \approx 15 \mu\text{m}$ (black dashed line) to $\lambda \approx 27 \mu\text{m}$ (red dashed line). Left: decay length (λ) versus a parameter b that captures monotonic and continuous changes in k_{on} , k_{off} and D_0 as shown in the right. Right: Variations in k_{on} , k_{off} and D_0 with b as defined by the equations shown in the plot. Black and red dashed lines indicate initial (small discs) and final (large discs) values for k_{on} , k_{off} and D_0 . **b**, Top: expression for the ratio of the recycling to the unbound module (λ_r^2/λ_u^2 , see Fig. 1c). Bottom: Sets of parameter values (clouds of points) compatible with all the assays considered in this report in the $(k_{\text{on}}, k_{\text{off}})$ plane. Isolines for $(\lambda_r^2/\lambda_u^2$ are also shown (see look-up table). The three conditions considered in this work are shown: large discs (average posterior length $l = 144 \mu\text{m}$ in the dataset; left), small discs (average $l = 80 \mu\text{m}$; centre) and *pent²* discs (right). These isolines convey the relative importance of the recycling and the unbound modules to the Dpp transport. **c**, PMAD scaling analysis for control and *pentagone* mutants. Left, Decay length λ of PMad gradients plotted as a function of posterior compartment length l . Raw and binned data (Bar, s.e.m) are shown together with a linear regression to the raw data. Right, bar plots showing the slopes ϕ of corresponding linear regressions for control (blue) and *pentagone* mutant experimental conditions (red). Number of biologically independent samples: $n = 45$ (control) and $n = 25$ (*pent²*). ***p-value < 0.00001; two-tailed two sample t-test with unequal variances. Bars, confidence intervals at 95%. **d**, UAS-GFP-Pentagone expression driven by ap-Gal4. In the right, higher magnification of the area boxed in the image to the left. Scale bars, 10 μm . **e**, GFP-Pentagone gradient profile in the ventral compartment. The profile is fitted to an exponential function (red) to determine the decay length shown. x , distance from the dorso-ventral boundary. **f**, eGFP-Dpp^{LOP} scaling analysis for control and *pentagone* mutants. Left, Decay length λ of eGFP-Dpp gradients plotted as a function of posterior compartment length l . Raw and binned data (Bar, s.e.m) are shown together with a linear regression to the raw data. Right, bar plot showing the slopes ϕ of corresponding linear regressions from these plots. Control experimental condition (blue) compared to *pentagone* mutant experimental condition (red). Number of biologically independent samples: $n = 157$ (control) and $n = 63$ (*pent²*). ***p-value < 0.00001; two-tailed two sample t-test with

unequal variances. Bars, confidence intervals at 95%. **g**, Sets of parameter values satisfying the constraints given by all the experimental assays represented in $(k_{\text{on}}, k_{\text{off}})$, (k_{on}, D_0) and (k, k_{off}) planes in the four experimental conditions: eGFP-Dpp^{LOP}-expressing discs of 144 μm and 80 μm average posterior length and *pent*² mutant discs of 130 μm and 85 μm average posterior length. **h**, Stacked bar chart showing the relative contribution of the different modules to λ^2 (described in Fig. 1e,f) in the four experimental conditions in e compared to the theoretical values of parameters in the extracellular diffusion (ExD) and transcytosis regimes of transport (Tr). **i**, Average extracellular fraction in control discs of 144 μm and 80 μm average posterior length and *pent*² mutant discs of 130 μm and 85 μm average posterior length. Box plot represents the minimum and the maximum, median, 25th and 75th percentile. n, number of biologically independent samples. **j**, Confocal images of PentGFP from the endogenous gene in discs of different sizes. Scale bar, 10 μm . Dotted lines, contour of discs. **k**, PentGFP average intensity in its expression domain as a function of the squared posterior length of the wing disc; Black, binned data. Orange dots, raw data. Bars, s.e.m. Vertical boxes indicate posterior width sizes $l = 144 \mu\text{m}$ (orange) and $l = 80 \mu\text{m}$ (blue).

[Source data](#)

[Extended Data Fig. 8 Gradient scaling by recycling: HSPGs.](#)

a, b, Scaling analysis for control and *dally* mutants. Left, decay length λ of eGFP-Dpp (**a**) and PMad gradients (**b**) plotted as a function of posterior compartment length l . Raw and binned data (bars, s.e.m.) are shown together with a linear regression to the raw data. Right, bar plots showing the slopes ϕ of corresponding linear regressions for control experimental conditions (blue) compared to *dally* mutant experimental conditions (red). Number of biologically independent samples: n = 93 (control) and n = 39 (*dally*^{gem}) (**a**); n = 43 (control) and n = 36 (*dally*^{gem}) (**b**). ***p-value < 0.00001; two-tailed two sample t-test with unequal variances. Bars, confidence intervals at 95%. **c**, Sets of parameter values satisfying the constraints given by all the experimental assays represented in $(k_{\text{on}}, k_{\text{off}})$, (k_{on}, D_0) and (k, k_{off}) planes in the four experimental conditions: eGFP-Dpp

discs of 144 μm and 80 μm average posterior length, *pent*² (average length, 130 μm) mutant and *dally^{gem}* mutant discs (average length, 174 μm). **d**, Stacked bar chart showing the relative contribution of the different modules to λ^2 (described in Fig. 1e,f) in the four experimental conditions compared to the theoretical values of parameters in the extracellular diffusion (ExD) and transcytosis regimes of transport (Tr). **e**, GBP-Alexa555 signal intensity as a function of time in discs expressing eGFP-Dpp^{Gal4} in control discs (left), *dally^{gem}* mutant discs (middle) and control discs following treatment with PI-PLC for 1h (right). Lines, fits to the phenomenological equation describing the internalized signal intensity dynamics $C_T(t)$. **f**, Values of k_N , k_r and k_0 estimated by the nanobody uptake assay in control discs, *dally^{gem}* mutant discs and PI-PLC treated discs expressing eGFP-Dpp^{Gal4}. **g**, Internalized GBP-Alexa555 fluorescence as a function of time in discs expressing eGFP-Dpp^{CRISPR} (control), discs expressing eGFP-Dpp^{CRISPR} and sflRNAi (sflRNAi) and control discs (no GFP-Dpp). Number of biologically independent samples: n = 3 for each condition. Data represented as the average curve. Shaded area, s.e.m. **h**, **i**, Confocal images of eGFP-Dpp^{CRISPR} (left) and internalized GBP-Alexa555 (right) after 85 min of incubation with the nanobody in control discs (**h**) and discs expressing sflRNAi in the posterior compartment (**i**). Posterior compartment, to the right from the GFP-Dpp source boundary. **j**, Decay length of the eGFP-Dpp^{CRISPR} gradient λ as a function of the posterior compartment width l . Red line, linear regression to the raw data. bars, s.e.m. eGFP-Dpp^{CRISPR} was visualized by means of a nanobody uptake assay (Methods). Number of biologically independent samples n = 38. **k**, Slope ϕ of the linear regressions for scaling plots corresponding to eGFP-Dpp^{LOP} (LOP) and eGFP-Dpp^{CRISPR} (CRISPR). Bars, confidence intervals of the fitted slope. **l**, Confocal images of photoconverted GBP-Dendra2* in eGFP-Dpp^{CRISPR}-expressing discs at different times after photoconversion (post-conversion). Before photoconversion, discs were incubated in GBP-Dendra2* solution for 45 min and extracellular GBP-Dendra2 was removed by an acid wash, so that only internalized GBP-Dendra2 is remaining. PhotoconvOgradient outside of the photoconverted region. **m**, The values of k_N , k_r and k_0 estimated by the nanobody uptake parameters for large discs expressing eGFP-Dpp^{CRISPR} versus eGFP-Dpp^{LOP}. Bars, confidence

intervals of the fits. Number of biologically independent samples $n = 10$ (eGFP-Dpp^{CRISPR}) and $n = 13$ (eGFP-Dpp^{LOP}). Scale bar, 10 μm (**h, l**).

[Source data](#)

Extended Data Table 1 Rates of Dpp transport

Extended Data Table 2 Parameterization assays

Supplementary information

[Supplementary Information](#)

This file contains the Supplementary Methods, Supplementary Notes, Supplementary Discussion, Supplementary Fig. 1, Supplementary Tables 1–2 and Supplementary References.

[Reporting Summary](#)

[Peer Review File](#)

[41586_2021_4346_MOESM4_ESM.avi](#)

Supplementary Video 1 **Photoconversion assay**. Left, a movie combining sequentially first, a confocal image of eGFP-Dpp^{LOP}, then an image of photoconverted endosomal GBP–Dendra2 (Dendra2*) before (pre-conversion) and finally images of Dendra2* at indicated times after photoconversion (post-conversion). Before conversion, following pulse-chase and acid wash, only internalized GBP–Dendra2 remains.

Photoconversion, to the left of the red dotted line. Note build-up of a Dendra2* gradient outside the photoconverted region. Right, average spatial distribution of GBP–Dendra2* fluorescence signal at indicated times after photoconversion. Shaded areas, s.e.m.; ($n = 7$ independent experiments).

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**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Romanova-Michaelides, M., Hadjivasiliou, Z., Aguilar-Hidalgo, D. *et al.* Morphogen gradient scaling by recycling of intracellular Dpp. *Nature* **602**,

287–293 (2022). <https://doi.org/10.1038/s41586-021-04346-w>

- Received: 23 April 2020
- Accepted: 10 December 2021
- Published: 22 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04346-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/s41586-021-04346-w>

- Article
- [Published: 24 November 2021](#)

The N501Y spike substitution enhances SARS-CoV-2 infection and transmission

- [Yang Liu](#)^{1,2} ✉nal,
- [Jianying Liu](#)^{2,3,4} ✉nal,
- [Kenneth S. Plante](#) [ORCID: orcid.org/0000-0002-7485-3167](#)^{2,3,4} ✉nal,
- [Jessica A. Plante](#) [ORCID: orcid.org/0000-0002-4768-7458](#)^{2,3,4},
- [Xuping Xie](#) [ORCID: orcid.org/0000-0003-0918-016X](#)¹,
- [Xianwen Zhang](#)¹,
- [Zhiqiang Ku](#) [ORCID: orcid.org/0000-0001-8347-0013](#)⁵,
- [Zhiqiang An](#) [ORCID: orcid.org/0000-0001-9309-2335](#)⁵,
- [Dionna Scharton](#)^{2,3,4},
- [Craig Schindewolf](#)^{2,4},
- [Steven G. Widen](#) [ORCID: orcid.org/0000-0002-6512-363X](#)¹,
- [Vineet D. Menachery](#) [ORCID: orcid.org/0000-0001-8803-7606](#)^{2,4},
- [Pei-Yong Shi](#) [ORCID: orcid.org/0000-0001-5553-1616](#)^{1,2} &
- [Scott C. Weaver](#) [ORCID: orcid.org/0000-0001-8016-8556](#)^{2,3,4}

[Nature](#) volume 602, pages 294–299 (2022)

- 34k Accesses
- 7 Citations
- 317 Altmetric
- [Metrics details](#)

Subjects

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

Abstract

The B.1.1.7 variant (also known as Alpha) of SARS-CoV-2, the cause of the COVID-19 pandemic, emerged in the UK in the summer of 2020. The prevalence of this variant increased rapidly owing to an increase in infection and/or transmission efficiency¹. The Alpha variant contains 19 nonsynonymous mutations across its viral genome, including 8 substitutions or deletions in the spike protein that interacts with cellular receptors to mediate infection and tropism. Here, using a reverse genetics approach, we show that of the 8 individual spike protein substitutions, only N501Y resulted in consistent fitness gains for replication in the upper airway in a hamster model as well as in primary human airway epithelial cells. The N501Y substitution recapitulated the enhanced viral transmission phenotype of the eight mutations in the Alpha spike protein, suggesting that it is a major determinant of the increased transmission of the Alpha variant. Mechanistically, the N501Y substitution increased the affinity of the viral spike protein for cellular receptors. As suggested by its convergent evolution in Brazil, South Africa and elsewhere^{2,3}, our results indicate that N501Y substitution is an adaptive spike mutation of major concern.

[Download PDF](#)

Main

Since its emergence in 2019⁴, more than 227 million SARS-CoV-2 infections have occurred worldwide, with more than 4.6 million fatalities. Despite exhibiting proofreading activity⁵, coronaviruses mutate frequently⁶, leading to thousands of mutations among the different strains. The most crucial mutations have involved the spike protein, which mediates binding to the angiotensin-converting enzyme 2 (ACE2) receptor via its S1 domain and membrane fusion via the S2 domain⁷. Spike substitutions influence host range, tissue tropism, transmission and pathogenesis⁸.

The first dominant spike substitution, D614G, enhances replication in human airway epithelial (HAE) cells and transmission in animal models^{9,10,11,12}. Additional spike mutations, including the N501Y mutation that occurred convergently in the UK, South Africa, Brazil and elsewhere, have been identified. The Alpha variant, which contains seven additional spike substitutions, emerged in September 2020 in the UK. It was estimated to be 70–80% more transmissible than its ancestor¹ and quickly spread worldwide^{2,3}. In vitro studies show enhanced replication of Alpha in human airway cells¹³. Although it is effectively neutralized by vaccine-induced antibodies^{14,15,16}, and is not consistently associated with more severe COVID-19¹⁷ disease, increased Alpha transmissibility could counteract increasing herd immunity, exacerbating the pandemic.

To investigate mechanisms of increased Alpha transmission, we used reverse genetics, a hamster model and HAE cultures to probe the effects of spike mutations. We determined the phenotypes of individual mutations as well as of the entire spike gene using competition fitness assays. Our results suggest that N501Y is a critical determinant of enhanced infection of the upper airway and transmission.

Results

To investigate phenotypes of individual Alpha spike substitutions, we infected hamsters intranasally with each mutant virus, as well as with chimeric viruses containing the entire spike gene (Extended Data Fig. [1a–c](#)). These mutations were engineered into the USA_WA1/2020 cDNA clone containing the dominant D614G spike substitution^{[10](#)} using site-directed mutagenesis^{[18](#)} (resulting in USA_WA1/2020-G614, hereafter referred to as the wild-type (WT) virus). These mutations did not affect Vero cell plaque phenotypes (Extended Data Fig. [1d](#)) and the RNA:plaque-forming units (PFU) ratios were similar to those for the virus expressing wild-type spike (Extended Data Fig. [1e](#)). We tested fitness of viruses expressing the mutant and Alpha spike proteins in an in vivo competition assay by mixing WT viruses expressing mutant and wild-type spike at a PFU ratio of 1:1 for intranasal infection of hamsters (Fig. [1a](#)). When assessed by quantification of viral RNA, these ratios were in a range of 0.72–1.54, presumably owing to minor variations in RNA:PFU ratios (Supplementary Table [1](#)). Although the enrichment of N501Y mutants after competition overlapped with this range in some cases, the increases in viral RNA and typically larger changes in PFU ratio indicated robust fitness advantages. This competition approach has been used extensively to study microbial fitness^{[19,20,21,22](#)} including for SARS-CoV-2^{[10,13](#)}, owing to its precision and reproducibility^{[23](#)}. Major advantages of this approach include the internally controlled replication of each virus, which eliminates host-to-host variation that can reduce experimental power, and the ability to assay strain ratios with more precision than when using individual virus titres. To verify the accuracy of estimates across a range of strain ratios, we performed experiments with each mixture using Sanger sequencing, and found a high degree of consistency between PFU- and genome-based estimates (Extended Data Fig. [2](#)). We also evaluated all virus stocks used for experiments by next-generation sequencing to confirm that there were no single-nucleotide polymorphisms (SNPs) with a frequency above 3%; sequencing after two serial passages in hamsters did not detect additional SNPs. These results indicate that the mutant mixtures were genetically stable and that the PFU ratios were robust.

Fig. 1: The screening of the SARS-CoV-2 Alpha variant spike substitutions in hamsters by competition assay.

 **figure 1**

a, Design of the hamster competition fitness studies. The mutant viruses were mixed 1:1 (PFU ratio) with WT virus and inoculated into donor hamsters intranasally at a total titre of 10^5 PFU per hamster. The donor hamsters were co-housed with recipient hamsters 1 day after infection. After 8 h of contact, the donors were removed. All hamsters were subjected to nasal washes daily until 4 days after infection and organs

were collected 4 days after inoculation or contact. **b–g**, Competition of different Alpha mutants with WT virus. **b, d, f**, Final:inoculum PFU ratio of WT virus expressing 8 individual spike mutations or the Alpha spike protein that includes all 8 mutations: nasal washes (**b**), tracheae (**d**) and lungs (**f**) in donor hamsters 4 days after inoculation. **c, e, g**, Final:donor inoculum ratios of WT virus expressing 8 individual spike mutations or the Alpha spike protein: nasal washes (**c**), tracheae (**e**) and lungs (**g**) in recipient hamsters 4 days after contact. In **b–g**, red dots represent individual hamsters ($n = 5$). In catseye plots, the horizontal line shows the mean and the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

Source data

Hamsters were sampled up to four days after infection, following peak viral loads¹⁰, using nasal washes and necropsied tracheae and lungs (Fig. [1a](#)). One day after infection, hamsters were co-housed with recipients for 8 h to assess transmission; recipients were maintained for an additional 4 days for equivalent sampling. Inocula and hamster samples were analysed by PCR with reverse transcription (RT–PCR) spanning the mutations, and the amplicons were analysed by Sanger sequencing and electropherogram peak analyses to determine mutant:wild-type ratios. Ratio changes during infection reflected the relative fitness of the mutants.

Competition results from donor nasal washes indicated that of the 8 mutations examined, only the deletion of codons 69 and 70 ($\Delta 69\text{--}70$), N501Y and the combined Alpha spike mutations conferred significant fitness advantages, with other substitutions showing inconsistently reduced fitness or no difference from the wild-type spike (Fig. [1b](#)). The same fitness trends were observed in recipient hamsters (Fig. [1c](#)). Examination of donor tracheae showed similar results, except that $\Delta 69\text{--}70$ had no effect on fitness (Fig. [1d](#)), whereas A570D increased fitness and S982A and D1118H decreased fitness. Following transmission, the same overall effects on fitness were maintained in recipient tracheae (Fig. [1e](#)), consistent with results from nasal washes (Fig. [1c](#)). Results from day-4 lungs were distinct from those in the upper airway (nasal washes and tracheae), with only slight fitness effects of the spike mutations; even the combined Alpha-spike mutations did not increase fitness significantly with respect to replication in donor lungs (Fig. [1f](#)). However, there were significant fitness gains for N501Y, A570D and Alpha-spike mutations in recipients, suggesting more efficient transmission, whereas S982A decreased fitness in recipients (Fig. [1g](#)). We also found that N501Y conferred an advantage across days 1–4 of nasal shedding, beginning on day 3 in donors, that was maintained throughout days 1–4 in recipients (Fig. [2a,b](#)). Among the other mutations, only $\Delta 69\text{--}70$ and the Alpha-spike mutations showed consistent fitness gains (Extended Data Fig. [3](#)).

Fig. 2: The SARS-CoV-2 spike N501Y mutant has a consistent advantage over wild-type spike in upper airway replication in hamsters.

 figure 2

a, b, Competition assays between N501Y and wild-type (WT) spike assessed by sampling nasal washes of donor (**a**) and recipient (**b**) hamsters on days 1 to 4 after

inoculation (donors) or contact (recipients). **c–f**, Competition assays between N501Y and wild-type spike in the tracheae and lungs of donor (**c, d**) and recipient hamsters (**e, f**) on days 2 and 4 after inoculation (donors) or contact (recipients). **g–j**, Competition assays between Alpha-FL virus and WT virus inoculated intranasally into hamsters. Nasal washes were taken on days 1 to 4 after inoculation (donors) (**g**) or contact (recipients) (**h**) and tracheae and lungs of donor (**i**) and recipient hamsters (**j**) 4 days after inoculation (donors) or contact (recipients). **k–n**, Competition between Alpha-FL and Alpha (WT spike) assessed by sampling nasal washes on days 1 to 4 after inoculation (donors) (**k**) or contact (recipients) (**l**) and tracheae and lungs of donor (**m**) and recipient hamsters (**n**) 4 days after inoculation (donors) or contact (recipients). The fitness advantages of the N501Y substitution and Alpha-FL compared with the wild-type both during infection (in donors) and after transmission of the virus to recipients are shown by the changes in ratios between the collected samples and inocula. Red dots represent individual hamsters (all $n = 5$, except $n = 10$ for day-1 nasal wash samples of N501Y group). In catseye plots, the horizontal line represents the mean, the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

Source data

Because it conferred the clearest phenotype among spike mutations, we examined N501Y in trachea and lungs sampled on days 2 and 4. Similar to nasal washes, recipient samples showed a consistent fitness advantage for N501Y at both timepoints, but no difference was detected in donor lungs (Fig. [2c–f](#)). Together, these results demonstrated a clear fitness advantage for N501Y in terms of upper airway shedding during days 1–4, including after intra-cage transmission.

To further ensure that the Sanger sequencing method did not introduce errors in our estimates of mutant ratios, all experiments with N501Y were repeated using next-generation sequencing. As shown in Extended Data Fig. [4](#), results using next-generation sequencing were very similar and statistically indistinguishable from those obtained with Sanger sequencing.

To examine whether the spike protein was the main determinant of Alpha replication and transmission, we generated a full-length Alpha infectious clone (Alpha-FL) (EPI_ISL_999340) incorporating all Alpha mutations, as well as chimeric viruses in which the spike was swapped (Extended Data Fig. [1c](#)). In competition assays with WT virus, Alpha-FL had a marked fitness advantage in nasal washes from donor hamsters on days 1 and 2, when titres were maximal (Extended Data Fig. [5](#)); similar results were observed in recipient hamsters (Fig. [2g, h](#)). Alpha-FL showed increased fitness in the trachea and lungs of both donors and recipients on day 4, after titres had peaked (Fig. [2i, j](#), Extended Data Fig. [5](#)). This difference in fitness of Alpha at different sites

over time could reflect the complex effects of innate host defences and/or lower stability in the nasal cavity. In competition against Alpha expressing wild-type spike (Alpha (WT spike)), Alpha-FL showed a consistent advantage on days 1–4 in nasal washes of both donor and recipient hamsters (Fig. 2k, l) and on day 4 in tracheae and lungs (Fig. 2m, n). This again indicated that the Alpha spike was a major fitness determinant during infection. Alpha-FL showed consistently reduced fitness in competition assays with Alpha (WT spike) in nasal washes of donors and recipients, and in trachea and lungs (Extended Data Fig. 6), again suggesting that some non-spike Alpha genome regions confer a fitness disadvantage *in vivo*.

We also assayed relative fitness of WT virus expressing N501Y or Alpha spike in recipient hamsters using the PFU ratio in donor hamster day-1 nasal wash (representing the transmitted population) as the starting PFU ratio rather than that of the original donor inoculum, with little effect on fitness estimates (Extended Data Fig. 7).

To examine the fitness effect of the Alpha spike substitutions in human cells, we used HAE and other human cell cultures (Fig. 3a). Replication of WT virus expressing N501Y and complete Alpha-spike mutants was significantly faster than that of WT virus at early stages of infection, as measured by PFU (Fig. 3b), but there was little difference in the number of RNA copies (Fig. 3c). The RNA:PFU ratios were significantly lower for virus expressing N501Y and Alpha spike compared with wild-type spike on days 1–3 (Fig. 3d), suggesting that viruses with these mutant spike proteins show increased specific infectivity in HAE cultures. In competition infection assays, N501Y showed significantly higher fitness at nearly all time points in Vero E6 and Calu-3 cells (Fig. 3e, f). In HAE cells, N501Y had a fitness advantage over wild-type spike and Alpha-FL spike had a fitness advantage over Alpha (WT spike) (Fig. 3g, h, Extended Data Fig. 8). Alpha-FL also showed a consistent fitness advantage over Alpha with wild-type spike, suggesting that, *in vitro*, the Alpha spike is the major fitness determinant of the Alpha variant (Fig. 3i). Overall, these *in vivo* and *in vitro* results suggest that N501Y is the major, but not the only, determinant of fitness of the Alpha variant.

Fig. 3: The spike N501Y mutation enhances viral replication in primary human airway cells and confers an advantage in competition with wild-type spike *in vitro*.

 **figure 3**

a, Experimental scheme. Viruses expressing wild-type spike, and N501Y and Alpha spike mutants were inoculated onto HAE cells at a multiplicity of infection (MOI) of 5 PFU per cell. After a 2-h incubation, the culture was washed with Dulbecco's PBS 3 times and maintained for 5 days. The secreted viruses were collected in Dulbecco's

phosphate-buffered saline (DPBS) by 30 min incubation at 37 °C every day. **b–d**, Viral replication kinetics and genomic RNA:PFU ratios. The amounts of infectious virus (**b**) and genomic RNA (**c**) were quantified by plaque assay and quantitative PCR with reverse transcription (RT–qPCR), respectively. The genomic RNA:PFU ratio (**d**) was calculated as an indication of virion infectivity. Dots represent individual biological replicates ($n = 6$) pooled from 2 independent experiments. Data are mean \pm s.e.m. Non-parametric two-tailed Mann–Whitney test with Bonferroni correction to account for multiple comparisons. Differences were considered significant if $P < 0.025$. **e–g**, Competition assay between WT virus expressing spike N501Y and WT virus on Vero E6 (**e**), Calu-3 (**f**) and HAE (**g**) cells. **h, i**, Competition assay between Alpha-FL and WT virus (**h**) and Alpha-FL and Alpha virus expressing wild-type spike protein (**i**) on HAE cells. **e–i**, The Vero E6, Calu-3 and HAE cells were infected with mixed viruses at MOI of 0.01, 0.1 and 5, respectively. Red dots represent individual cell cultures ($n = 6$), pooled from two independent experiments. In catseye plots, the horizontal line represents the mean, the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

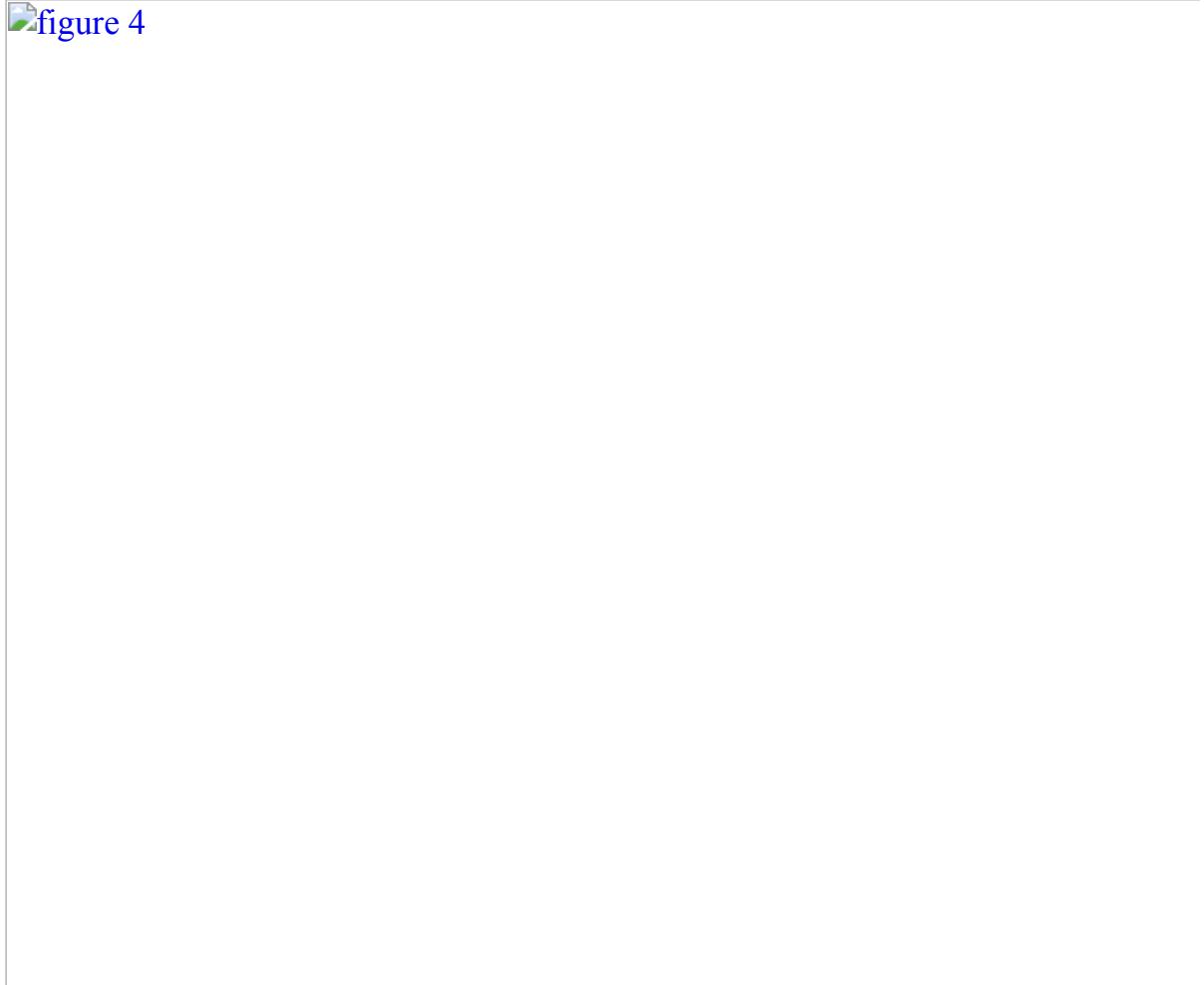
Source data

To assess virulence, we examined hamsters infected with each mutant virus individually for weight loss; we found no significant differences between wild-type, N501Y and Alpha spike proteins on weight loss (Extended Data Fig. [9a, b](#)). However, the trend of less weight loss in hamsters infected with WT virus suggested that differences may emerge in experiments with a larger cohort. Nasal washes showed consistently higher titres on days 1–3 and 5 for both N501Y and Alpha spike, but only when measured in PFU. These differences were significant only on day 1 (Extended Data Fig. [9c, d](#)), reflecting the lower sensitivity of individual infections compared with competition assays. Genomic RNA:PFU ratios were higher for virus with wild-type spike than for mutant spike (Extended Data Fig. [9e](#)), again suggesting that the viruses with mutant spike have greater specific infectivity. Additional studies are needed to determine whether these specific infectivity differences are biologically relevant. Trachea and lung samples showed no significant differences in titre when measured as PFU or number of copies of viral RNA (Extended Data Fig. [9f–h](#)). Together, these results again suggest that N501Y results in increased viral fitness in the hamster upper airway.

As the SARS-CoV-2 pandemic has progressed, several mutations have occurred convergently in multiple variants; of the 8 Alpha spike substitutions, only Δ69–70 and N501Y have evolved convergently, consistent with our fitness findings (Extended Data Table [1](#)). The frequency of N501Y worldwide increased markedly after October 2020 (Fig. [4a](#)). These data corroborate our results indicating that the N501Y substitution confers a major fitness advantage.

Fig. 4: The spike N501Y substitution spread quickly and increases spike protein binding affinity for the human ACE2 receptor.

 figure 4



a, The frequency of the N501Y substitution over time in all genomic SARS-CoV-2 sequences available from the GISAID database worldwide up to May 2021. The blue bars represent the total numbers of SARS-CoV-2 genomes sequenced worldwide. The red line indicates the percentage of N501Y variant in total SARS-CoV-2 genomes. **b**, The predicted binding site of spike N501 and Y501 residues on the human ACE2 receptor. **c, d**, Binding affinities of wild-type spike (**c**) and spike(N501Y) (**d**) to the human ACE2 receptor. K_D , dissociation constant; k_{off} , dissociation rate constant; k_{on} , association rate constant. The affinity of ACE2 to the N501Y mutant RBD is below the detection limit and is plotted as $<10^{-12}$. Data are derived from a single experiment.

[Source data](#)

We next examined the mechanism behind the fitness advantage of N501Y, focusing on its effect on receptor binding. N501Y is located in the receptor-binding domain (RBD) of spike (Fig. 4b) and is proposed to result in increased interactions with the human

ACE2 receptor at K353, increasing the affinity for the receptor. To test this, we performed binding assays using recombinant spike RBD and human ACE2 proteins on a bio-layer interferometry system (Fig. 4c, d). The N501Y substitution increased binding, as indicated by the more than 530-fold improvement in K_D and more than 819-fold improvement in k_{off} . These results are consistent with earlier work demonstrating that spike residue 501 is a critical residue for ACE2 receptor binding²⁴ and suggest that N501Y increases viral fitness for replication in the upper airway, resulting in enhanced transmission, via enhanced spike–receptor interactions.

Discussion

To corroborate epidemiological findings of more efficient transmission of the Alpha variant¹ of SARS-CoV-2, and to identify genetic determinants of this phenotype, we performed experimental infections of hamsters and HAE cells with each of the eight individual spike gene mutations found in Alpha. Of these mutations, only N501Y and the deletion of codons 69 and 70 resulted in consistent fitness advantages for replication in the upper airway of hamsters (with increased shedding in nasal secretions) as well as in human cells. In most experiments, N501Y alone had a phenotype similar to that of the chimeric WT virus containing the complete Alpha spike or the complete Alpha-FL, suggesting that N501Y is the major determinant driving increased transmission. Finally, our studies of chimeric SARS-CoV-2 with swapped spike genes suggest—as in another recent study²⁵—that non-spike genomic regions may interact in vivo with innate immune responses to both positively and negatively affect replication in a site- and time-dependent manner.

Although the data are not directly comparable, the typical relative fitness values of 2–3 that we observed with the N501Y substitution in the hamster or HAE models are consistent with reproduction numbers between 1.4 and 1.8 estimated by epidemiological modelling²⁶. This substitution, which is also present in other variants as a result of convergent evolution, is therefore a major adaptive spike mutation of concern. The N501Y mutation also enables SARS-CoV-2 infection of mice²⁷. Its continued evolution in human viral strains only in combination with other spike mutations suggests epistatic interactions. The mechanism of enhanced N501Y fitness is likely to be related to increased replication and shedding in the nasal cavity, leading to more efficient airborne and/or fomite-mediated transmission, both of which are important for the spread of SARS-CoV-2. More detailed transmission experiments are needed to discriminate between these two modes. In addition to upper airway shedding, increased transmission could involve differences in virus stability and infectivity (for example, the 50% infectious dose).

Tests to date with N501Y indicate that this mutation results in minimal changes in susceptibility to immunity, reflected in neutralizing antibody titre (PRNT₅₀)

values^{14,15,16}, suggesting that it does not increase resistance to vaccine protection. These data along with ours suggest that selection of Alpha in nature has occurred owing to more efficient receptor binding rather than immune escape.

Although Δ69–70 has also been found in a mink variant as well as in wild-type SARS-CoV-2 strains (Extended Data Table 1), it has not evolved convergently in major variants. Furthermore, in the USA, despite its presence since October, 2020, it has not increased substantially in frequency²⁸. These findings corroborate our results indicating that Δ69–70 has less effect on transmission than N501Y. However, the frequent co-occurrence of this deletion²⁹ with N501Y suggests possible epistasis. Furthermore, initial epidemiologic estimates from the UK have indicated that N501Y alone increased transmissibility by about 10%, whereas strains with Δ69–70 showed 70–80% higher transmissibility¹, suggesting that the deletion is also important for Alpha transmission.

Considering that only two of the eight spike substitutions conferred consistent fitness effects on infection of the upper airway and transmission, the origins and roles of the other six mutations deserve consideration. The first of these to appear were deletions of codons 69–70 and 145, followed by all the other except for S982A by March 2020 (Supplementary Table 2). However, from September 2020, all eight mutations—and especially N501Y—began increasing in prevalence (Fig. 4a, Supplementary Table 2). Our data indicate that of these only Δ69–70 and N501Y have a consistent phenotype, and some of the other mutations reduced fitness in certain experiments. This suggests that these other mutations occurred through drift mechanisms such as founder and hitchhiking effects, and were maintained by linkage to Δ69–70 and N501Y, or by recombination, which occurs frequently during SARS-CoV-2 replication⁶. Other possibilities include selection by convalescent plasma or antiviral treatments during chronic infections. It is also likely that mutations in other parts of the genome contribute to the fitness of the Alpha variant, as indicated by our results from chimeric virus.

The hamster is a useful model for SARS-CoV-2 infection in many respects¹⁰, but its ACE2 receptors differ slightly the human receptors³⁰. The spike binding studies should be repeated with the hamster receptors to confirm our affinity findings. Although we used infection of primary HAE cells to address this limitation, the use of a primate model would further substantiate our in vivo findings.

In conclusion, we used authentic SARS-CoV-2 to demonstrate that only two out of eight spike substitutions in the Alpha variant are major contributors its enhanced transmission, and the N501Y substitution appears to be the major determinant. This

mutation is present in several global regions and should be closely monitored to anticipate public health measures needed to control the spread of SARS-CoV-2.

Methods

Ethics statement

Hamster studies were performed in accordance with the guidance for the Care and Use of Laboratory Animals of the University of Texas Medical Branch (UTMB). The protocol was approved by the Institutional Animal Care and Use Committee at UTMB. All the hamster operations were performed under anaesthesia by isoflurane to minimize animal suffering.

Animals and Cells

The Syrian golden hamsters (HsdHan:AURA strain) were purchased from Envigo. African green monkey kidney epithelial Vero E6 cells (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco/Thermo Fisher) with 10% fetal bovine serum (FBS; HyClone Laboratories) and 1% antibiotic/streptomycin (Gibco). Human lung adenocarcinoma epithelial Calu-3 cells (ATCC) were maintained in a high-glucose DMEM containing sodium pyruvate and GlutaMAX (Gibco) with 10% FBS and 1% penicillin/streptomycin at 37 °C with 5% CO₂. The EpiAirway system is a primary human airway 3D tissue model purchased from MatTek Life Science (Ashland). This EpiAirway system was maintained with the provided culture medium at 37 °C with 5% CO₂ following the manufacturer's instruction. All other culture medium and supplements were purchased from Thermo Fisher Scientific. All cell lines were verified and tested negative for mycoplasma.

Generation of SARS-CoV-2 mutant viruses

Generation of individual point mutant viruses and WT virus expressing Alpha spike

Individual point mutations in the spike gene (Δ 69–70, Δ 145, N501Y, A570D, P681H, T716I, S982A and D1118H) or 8 spike mutations were introduced into a subclone pcc1-CoV-2-F5-7 containing the spike gene of SARS CoV-2-D614G¹⁷ by overlap fusion PCR.

Generation of Alpha-FL and Alpha-FL (WT spike)

The sequence for constructing Alpha-FL and Alpha-non-spike was downloaded from GISAID under accession ID EPI_ISL_999340. In brief, individual point mutations (NSP3: P153L, T183I, A890D, I1412T; NSP6: SGF106-108del; NSP12: P323L; Spike: HV69-70del, Y145del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H; ORF8: Q27stop, R52I, Y73C, S84L; N: D3L, R203K, G204R, S235F) were introduced into the three subclones pcc1-Alpha-F1-3, puc57-Alpha-F4 and pcc1-Alpha-F5-7 by overlap fusion PCR. To generate the Alpha-FL (WT spike), the spike gene of Alpha-FL was replaced with the spike gene of SARS CoV-2-D614G. All the primers used for the construction are listed in Supplementary Table 3.

The full-length infectious clones of the variant SARS-CoV-2 viruses were assembled by in vitro ligation of contiguous DNA fragments the protocol as previously described³¹. In vitro transcription was then preformed to synthesize full-length genomic RNA. For recovery of the mutant viruses, the RNA transcripts were electroporated into Vero E6 cells. The viruses from electroporated cells were collected at 40 h after electroporation and served as P0 stocks. All viruses were passaged once on Vero E6 cells for subsequent experiments and subjected to next-generation sequencing after RNA extraction to confirm the introduction and stability of substitutions. Viral titres were determined by plaque assay on Vero E6 cells. All virus preparation and experiments were performed in a biosafety level 3 facility. Viruses and plasmids are available from the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) at the UTMB.

RNA extraction, RT-PCR, RT-qPCR and Sanger sequencing

Cell culture supernatants or clarified tissue homogenates were mixed with a fivefold excess of TRIzol LS Reagent (Thermo Fisher Scientific). Viral RNAs were extracted according to the manufacturer's instructions. The extracted RNAs were dissolved in 20 µl nuclease-free water. For the validation of mutant viruses, two microlitres of RNA samples were used for reverse transcription by using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) with random hexamer primers. Nine DNA fragments flanking the entire viral genome were amplified by PCR. The resulting DNAs were cleaned up by the QIAquick PCR Purification Kit, and the genome sequences were determined by Sanger sequencing at GENEWIZ (South Plainfield).

To quantify mutant ratios for competition assays, RT-PCR products were amplified from extracted RNA using a SuperScript III One-Step RT-PCR kit (Invitrogen). A 20-µl reaction was assembled in PCR 8-tube strips through the addition of 10 µl 2× reaction mix, 0.4 µl SuperScript III RT/Platinum Taq Mix, 0.8 µl Forward Primer (10 µM), 0.8 µl reverse primer (10 µM), 2 µl RNA, and 6 µl Rnase-free water. Primers are listed in Supplementary Table 3. Reverse transcription and amplification were

completed using the following protocol: (1) 55 °C, 30 min; 94 °C, 2 min; (2) 94 °C, 15 s; 60 °C, 30 s; 68 °C, 1 min; 40 cycles; (3) 68 °C, 5 min; (4) indefinite hold at 4 °C.

The presence and size of the desired amplicon was verified with 2 µl of PCR product on an agarose gel. The remaining 18 µl were purified by a QIAquick PCR Purification kit (Qiagen) according to the manufacturer's protocol.

Sequences of the purified RT–PCR products were generated using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). The sequencing reactions were purified using a 96-well plate format (EdgeBio) and analysed on a 3500 Genetic Analyzer (Applied Biosystems). The peak electropherogram height representing each mutation site and the proportion of each competitor was analysed using the QSVanalyser program³².

Next generation sequencing

The competition results between N501Y and WT in hamsters and HAE cells derived via Sanger sequencing were confirmed using next generation sequencing methods. In brief, the viral RNAs from the N501Y–WT competition groups were used for a specific one-step RT–PCR that containing the A23063T mutation site, primers for the RT–PCR are listed in Supplementary Table 3. The PCR products were then purified by a QIAquick PCR Purification kit (Qiagen) according to the manufacturer's protocol. Dual-indexed adapter sequences (New England BioLabs) were added with 5 cycles of PCR. Samples were pooled and sequenced by the UTMB Next Generation Sequencing Core on an Illumina MiniSeq Mid-Output flow cell with the paired-end 150 base protocol. The reads were filtered for *Q*-scores of 37 at the A23063T mutation site and adjacent bases and counted. The input ratios and output ratios of the two viruses were obtained and used for the relative replicative fitness analysis.

Plaque assay

Approximately 1.2×10^6 Vero E6 cells were seeded to each well of 6-well plates and cultured at 37 °C, 5% CO₂ for 16 h. Virus was serially diluted in DMEM with 2% FBS and 200 µl diluted viruses were transferred onto the monolayers. The viruses were incubated with the cells at 37 °C with 5% CO₂ for 1 h. After the incubation, overlay medium was added to the infected cells per well. The overlay medium contained DMEM with 2% FBS, 1% penicillin/streptomycin and 1% sea-plaque agarose (Lonza). After a 2-day incubation, plates were stained with neutral red (Sigma-Aldrich) and plaques were counted on a light box. The detection limitation of the plaque assay was 10 PFU ml⁻¹.

Quantitative real-time RT–PCR assays

RNA copies of SARS-CoV-2 samples were detected by quantitative real-time RT-PCR (RT-qPCR) assays were performed using the iTaq SYBR Green One-Step Kit (Bio-Rad) on the QuantStudio 7 Flex system (Applied Biosystems) following the manufacturers' protocols. Primers are listed in Supplementary Table 3. The absolute quantification of viral RNA was determined by a standard curve method using an RNA standard (in vitro transcribed 3,480 bp containing genomic nucleotide positions 26,044 to 29,883 of SARS-CoV-2 genome).

Viral infection of cell lines

Approximately 3×10^5 Vero E6 or Calu-3 cells were seeded onto each well of 12-well plates and cultured at 37 °C, 5% CO₂ for 16 h. N501Y or Alpha-spike mutants were mixed with wild-type viruses and inoculated onto Vero E6 and Calu-3 cells at a MOI of 0.01 and 0.1, respectively. The mixed virus was incubated with the cells at 37 °C for 2 h. After infection, the cells were washed with DPBS 3 times to remove any unattached virus. One millilitre of culture medium was added into each well for the maintenance of the cells. At each time point, 100 µl of culture supernatants were lysed in TRIzol LS reagent for the detection of competition assay, and 100 µl of fresh medium was added into each well to replenish the culture volume. The cells were infected in triplicate for each group of viruses. All samples were stored at -80 °C until analysis.

Viral infection in a primary human airway cell culture model

The EpiAirway system is a primary human airway 3D mucociliary tissue model consisting of normal, human-derived tracheal/bronchial epithelial cells. For viral replication kinetics, wild-type, N501Y or Alpha-spike mutant viruses were inoculated onto the culture at a MOI of 5, respectively. After 2 h infection at 37 °C with 5% CO₂, the inoculum was removed, and the culture was washed 3 times with DPBS. The infected epithelial cells were maintained without any medium in the apical well, and medium was provided to the culture through the basal well. The infected cells were incubated at 37 °C, 5% CO₂. From 1–5 days, 300 µl of DPBS were added onto the apical side of the airway culture and incubated at 37 °C for 30 min to elute the released viruses. Titres of samples taken immediately after washing the inoculum were not plotted because initial experiments showed the presence of very little or no detectable virus. All virus samples in DPBS were stored at -80 °C.

Spike RBD and ACE2 binding

The human ACE2 protein was purchased from Sino Biological (catalogue (cat.) no. 10108-H08H) and the human IgG1 Fc-tagged RBD proteins were made in-house

as previously described³³. The affinity measurement was performed on the ForteBio Octet RED 96 system (Sartorius). In brief, wild-type or N501Y mutant RBD proteins ($20 \mu\text{g ml}^{-1}$) were captured onto protein A biosensors for 300 s. The loaded biosensors were then dipped into the kinetics buffer for 10 s for adjustment of baselines.

Subsequently, the biosensors were dipped into serially diluted (1.23~300 nM) human ACE2 protein for 200 s to record association kinetics and then dipped into kinetics buffer for 400 s to record dissociation kinetics. Kinetic buffer without ACE2 was used to correct the background. The Octet Data Acquisition 9.0 software was used to collect affinity data. For fitting of K_D values, Octet Data Analysis software V11.1 was used to fit the curve by a 1:1 binding model and use of the global fitting method.

Hamster infections

Four- to six-week-old male golden Syrian hamsters, strain HsdHan:AURA (Envigo), were randomly selected (although the investigators were not blinded to the assignments) and inoculated intranasally with $100 \mu\text{l}$ SARS-CoV-2. For transmission competition assays, 5 donor hamsters received a mixture containing 5×10^4 PFU of WT virus and an equal amount of variant virus. One day later, 1 infected donor hamster was co-housed with one naive hamster for 8 h (10 pairs for N501Y group, 5 pairs for the other groups) and the donors were returned to their cages. For virus replication assays, the hamsters received DMEM with 2% FBS and 1% penicillin/streptomycin (Mock, $n = 4$), WT virus ($n = 9$), N501Y mutant virus ($n = 9$), Alpha-spike mutant virus ($n = 9$) at a dose of 10^4 PFU per hamster. The infected hamsters were weighed and monitored for signs of illness daily. Nasal washes were collected in $400 \mu\text{l}$ sterile DPBS at indicated time points. For the transmission competition study, hamsters were humanely euthanized for organ collections at 2 or 4 days after inoculation (donors) or after contact (recipients). For the virus replication study, 4 hamsters in each group were euthanized for organ collection 2 days after infection and others were euthanized 7 days after infection. The collected tracheae and lungs were placed in a 2-ml homogenizer tube containing 1 ml of maintenance medium (DMEM supplemented with 2% FBS and 1% penicillin/streptomycin) and stored at -80°C . Samples were subsequently thawed, lung or tracheae were homogenized for 1 min at 26 s^{-1} , and debris was pelleted by centrifugation for 5 min at 16,100g. Infectious titres were determined by plaque assay. Genomic RNAs were quantified by RT-qPCR. Ratios of mutant:wild-type RNA were determined by RT-PCR with quantification of Sanger peak heights.

Competition assays

For the competition assay on Vero E6, Calu-3 and primary human epithelial airway (HAE) cells, WT and mutant viruses were mixed and inoculated onto the cells at a

final MOI of 0.01, 0.01 and 5, respectively. For the competition in hamsters, 100 µl mixtures of WT and variant viruses (total 1×10^5 PFU per hamster) were inoculated intranasally into 4–6-week-old Syrian hamsters. On 1–4 days after inoculation (donors) or after contact (recipients), the infected hamsters were sampled for competition detection. An aliquot of the inoculum used for both hamster and cell infections was back-titrated to confirm the initial ratio of viruses. All samples were stored in –80 °C freezer prior to analysis.

Validation of competition assay by Sanger sequencing

To validate the consistency and accuracy of competition assay by Sanger sequencing, the WT and mutant viruses were mixed at ratios of 10:1, 5:1, 3:1, 1:1, 1:3, 1:5 and 1:10 based on their PFU titres (total 10^6 PFU viruses) or mixed with 10^6 , 10^5 , 10^4 , 10^3 and 10^2 PFU of the two viruses at a ratio of 1:1. The total RNA of these mixed viruses was isolated and amplified by RT–PCR. The ratios of WT:mutant were calculated by the peak heights of Sanger sequencing. Data were analysed by linear regression with correlation coefficients (r) and significance (P).

Statistics

Male hamsters were randomly allocated into different groups. The investigators were not blinded to allocation during the experiments or to the outcome assessment. No statistical methods were used to predetermine sample size. Dead hamster were excluded from sample collections and data analysis. Descriptive statistics have been provided in the figure legends. For in vitro replication kinetics, Kruskal–Wallis analysis of variance was conducted to detect any significant variation among replicates. If no significant variation was detected, the results were pooled for further comparison. Differences between continuous variables were assessed with a two-tailed non-parametric Mann–Whitney test. The PFU, genomic copies and RNA:PFU ratios were analysed using non-transformed values. The weight loss data are shown as mean ± s.d. and statistically analysed using two-way ANOVA with Turkey’s multiple comparison. Analyses were performed in Prism version 9.0 (GraphPad).

For virus competition experiments, relative replicative fitness values for mutant virus compared to G614 WT virus were analysed according to $w = (f_0/i_0)$, where w is relative fitness, i_0 is the initial mutant:WT ratio and f_0 is the final mutant:WT ratio after competition. Sanger sequencing (initial timepoint T_0) counts for each virus strain being compared were based on average counts over three replicate samples of inocula per experiment, and post-infection (timepoint T_1) counts were taken from samples of individual subjects. For cell culture samples, multiple experiments were performed, so that f_0/i_0 was clustered by experiment. To model f_0/i_0 , the ratio T_0/T_1 was found

separately for each subject in each strain group, log (base 10) transformed to an improved approximation of normality and modelled by analysis of variance with relation to group, adjusting by experiment when appropriate to control for clustering within experiment. Specifically, the model was of the form $\log_{10}(\text{Ratio at } T_1/\text{Ratio at } T_0) \sim \text{Experiment} + \text{Group}$. Fitness ratios between the two groups (the model's estimate of $w = (f_0/i_0)$) were assessed per the coefficient of the model's Group term, which was transformed to the original scale as 10^{coefficient}. This modelling approach compensates for any correlation due to clustering within experiment similarly to that of corresponding mixed effect models and is effective since the number of experiments was small. Statistical analyses were performed using R statistical software (R Core Team, version 3.6.1). In all statistical tests, two-sided $\alpha = 0.05$. Catseye plots³⁴, which illustrate the normal distribution of the model-adjusted means, were produced using the catseyes package³⁵.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The sequences, strain dates and locations of collection of SARS-CoV-2 variants include USA_WA1/2020 ([MT020880](#) from the NCBI database) and the Alpha variant (EPI_ISL_999340 from the GISAID database). All other information is available upon request. [Source data](#) are provided with this paper.

References

1. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **26**, 2002106 (2021).
2. Galloway, S. E. et al. Emergence of SARS-CoV-2 B.1.1.7 lineage—United States, December 29, 2020–January 12, 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 95–99 (2021).
3. Claro, I. M. et al. Local transmission of SARS-CoV-2 lineage B.1.1.7, Brazil, December 2020. *Emerg. Infect. Dis.* **27**, 970–972 (2021).
4. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

5. Smith, E. C., Blanc, H., Surdel, M. C., Vignuzzi, M. & Denison, M. R. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.* **9**, e1003565 (2013).
6. Gribble, J. et al. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **17**, e1009226 (2021).
7. Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
8. 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 e1220 (2020).
9. Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
10. Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
11. Chan, J. F. et al. Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in a golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin. Infect. Dis.* **71**, 2428–2446 (2020).
12. Zhou, B. et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **592**, 122–127 (2021).
13. Touret, F. et al. Replicative fitness of a SARS-CoV-2 20I/501Y.V1 variant from lineage B.1.1.7 in human reconstituted bronchial epithelium. *mBio* **12**, e0085021 (2021).
14. 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).
15. 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).
16. Liu, Y. et al. Neutralizing activity of BNT162b2-elicited serum—preliminary report. *N. Engl. J. Med.* **384**, 1466–1468 (2021).

17. 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).
18. Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e3 (2020).
19. Wiser, M. J. & Lenski, R. E. A comparison of methods to measure fitness in *Escherichia coli*. *PLoS ONE* **10**, e0126210 (2015).
20. Grubaugh, N. D. et al. Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. *Cell Host Microbe* **19**, 481–492 (2016).
21. Bergren, N. A. et al. ‘Submergence’ of Western equine encephalitis virus: evidence of positive selection argues against genetic drift and fitness reductions. *PLoS Pathog.* **16**, e1008102 (2020).
22. Coffey, L. L. & Vignuzzi, M. Host alternation of chikungunya virus increases fitness while restricting population diversity and adaptability to novel selective pressures. *J. Virol.* **85**, 1025–1035 (2011).
23. Liu, J. et al. Role of mutational reversions and fitness restoration in Zika virus spread to the Americas. *Nat. Commun.* **12**, 595 (2021).
24. 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).
25. Thorne, L. G. et al. Evolution of enhanced innate immune evasion by the SARS-CoV-2 B.1.1.7 UK variant. Preprint at <https://doi.org/10.1101/2021.06.06.446826> (2021).
26. Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
27. Gu, H. et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603–1607 (2020).
28. Larsen, B. B. & Worobey, M. Identification of a novel SARS-CoV-2 spike 69–70 deletion lineage circulating in the United States. *virological.org* <https://virological.org/t/identification-of-a-novel-sars-cov-2-spike-69-70-deletion-lineage-circulating-in-the-united-states/577> (2020).

29. Kemp, S. A. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
30. Chan, J. F. et al. Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin. Infect. Dis.* **71**, 2428–2446 (2020).
31. Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e843 (2020).
32. Carr, I. M. et al. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics* **25**, 3244–3250 (2009).
33. Ku, Z. et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **12**, 469 (2021).
34. Cumming, G. The new statistics: why and how. *Psychol. Sci.* **25**, 7–29 (2014).
35. Andersen, C. Catseyes: create catseye plots illustrating the normal distribution of the means. R package version 0.2.3 (2019).

Acknowledgements

This research was supported by grants from the NIA and NIAID of the NIH (AI153602 and AG049042 to V.D.M., and R24AI120942 to S.C.W.); a STARs award provided by the University of Texas System to V.D.M.; a Welch Foundation grant AU-0042-20030616; and Cancer Prevention and Research Institute of Texas (CPRIT) grants RP150551 and RP190561 to Z.A. P.-Y.S. was supported by NIH grant AI134907 and UL1TR001439, and by awards from the Sealy & Smith, Kleberg, John S. Dunn, Amon G. Carter, Gilson Longenbaugh and Summerfield Robert foundations. J.L. was supported by a James W. McLaughlin Fellowship Fund.

Author information

Author notes

1. These authors contributed equally: Yang Liu, Jianying Liu, Kenneth S. Plante

Affiliations

1. Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

Yang Liu, Xuping Xie, Xianwen Zhang, Steven G. Widen & Pei-Yong Shi

2. Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX, USA

Yang Liu, Jianying Liu, Kenneth S. Plante, Jessica A. Plante, Dionna Scharton, Craig Schindewolf, Vineet D. Menachery, Pei-Yong Shi & Scott C. Weaver

3. World Reference Center for Emerging Viruses and Arboviruses, University of Texas Medical Branch, Galveston, TX, USA

Jianying Liu, Kenneth S. Plante, Jessica A. Plante, Dionna Scharton & Scott C. Weaver

4. Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, USA

Jianying Liu, Kenneth S. Plante, Jessica A. Plante, Dionna Scharton, Craig Schindewolf, Vineet D. Menachery & Scott C. Weaver

5. Texas Therapeutics Institute, Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA

Zhiqiang Ku & Zhiqiang An

Contributions

Conceptualization: Y.L., J.L., K.S.P., J.A.P., X.X., V.D.M., P.-Y.S. and S.C.W.

Methodology: Y.L., J.L., K.S.P., J.A.P., X.X., X.Z., Z.K., Z.A., D.S., C.S., S.G.W.,

V.D.M., P.-Y.S. and S.C.W. Investigation: Y.L., J.L., K.S.P., J.A.P., V.D.M., P.-Y.S. and

S.C.W. Resources: X.X., Z.K. and Z.A. Data curation: Y.L., J.L., K.S.P., J.A.P.,

V.D.M., P.-Y.S. and S.C.W. Writing, original draft: Y.L., P.-Y.S. and S.C.W. Writing,

review and editing: Y.L., J.L., K.S.P., J.A.P., V.D.M., P.-Y.S. and S.C.W. Supervision:

P.-Y.S. and S.C.W. Funding acquisition: Z.A., V.D.M., P.-Y.S. and S.C.W.

Corresponding authors

Correspondence to [Pei-Yong Shi](#) or [Scott C. Weaver](#).

Ethics declarations

Competing interests

X.X., V.D.M. and P.-Y.S. have filed a US provisional patent 63/000,713 'Reverse genetic system of SARS-CoV-2', filed on 27 March 2020, on the reverse genetic system and reporter SARS-CoV-2. The other authors declare no competing interests.

Peer review information

Nature thanks Stanley Perlman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 The construction and morphology of SARS-CoV-2 mutants.](#)

a, The reverse genetic construction design of all the individual and combined mutations on the wt backbone (USA_WA1/2020 spike D614G mutant). L, leader sequence; S, Spike gene; Open reading frames, ORFs; E, envelope glycoprotein gene; M, membrane glycoprotein gene; N, nucleocapsid gene; UTR, untranslated region. **b**, The location of all 8 Alpha B.1.1.7 substitutions and D614G on the SARS-CoV-2 spike protein trimer. **c**, The schematic designs of the chimeric SARS-CoV-2 viruses. The wt is the USA_WA1/2020 strain with the spike D614G substitution. The Alpha-spike/WT-non-spike contains the spike gene from the Alpha variant in the WT backbone. The WT-spike /Alpha-non-spike has the WT spike gene with all non-spike gene regions from the Alpha variant. The Alpha-FL is the infectious clone of full-length Alpha variant. **d**, The morphologies of all the rescued mutant SARS-CoV-2 variants. The plaques were stained 2.5 days post-infection of Vero E6 cells. **e**, The virus stocks of individual recombinant SARS-CoV-2 mutants, as well as Alpha-FL were quantified for their genomic RNA and infectious plaque-forming units by RT-qPCR and plaque assay, respectively. The genomes/PFU ratio was calculated to determine specific infectivity. Dots represent individual biological replicates from 4 different aliquots of viruses. The red dotted line indicates the average genome:PFU

ratio of WT. The values in the graph represent the mean \pm standard deviation. A two-tailed non-parametric Mann-Whitney test was used to determine significant differences between the variants and wt. P values were adjusted using the Bonferroni correction to account for multiple comparisons. No statistical differences were detected among groups.

Extended Data Fig. 2 Validation of SARS-CoV-2 ratio determination by Sanger sequencing.

a–l (Left panels), The correlation between input PFU ratios and output RT-PCR amplicon ratios determined by Sanger sequencing. WT and Mutant viruses were mixed at PFU ratios of 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, or 1:10. Total RNAs of the virus mixtures were extracted and amplified by RT-PCR. The WT/Mutant ratios were calculated by the peak heights of Sanger sequencing. Data were analyzed by linear regression with correlation coefficients (r) and significance (P). Symbols represent individual replicates, and the solid line represents the fitted line. **a–l (Right panels)**, Assay range evaluation. The ratio of wt/Mutant virus mixtures calculated from Sanger sequencing were consistent when using a wide range of virus amounts. The WT/Mutant viruses were mixed at 1:1 PFU ratio. The total titers of the mixed viruses were 10^6 , 10^5 , 10^4 , 10^3 , and 10^2 PFU. The total RNA of virus mixture was isolated and amplified by RT-PCR. The wt/Mutant ratios were calculated by the peak heights from Sanger sequencing. Symbols represent individual replicates, bar heights represent the mean, and error bars represent the standard deviation. **a–l**, Data are derived from a single experiment conducted in triplicate ($n = 3$).

Extended Data Fig. 3 The competition of other SARS-CoV-2 mutants with wt in hamsters.

a–h, Eight SARS-CoV-2 spike mutants: Δ 69–70 (**a**), Δ 145 (**b**), A570D (**c**), P681H (**d**), T716I (**e**), S982A (**f**), D1118H (**g**) and Alpha-spike/WT-non-spike (**h**) were mixed with wt virus at a PFU ratio of 1:1. The mixture was then inoculated intranasally into donor hamsters and transmitted to the recipient hamsters following the scheme in Figure 1a. The total titer for infection was 10^5 PFU per hamster. The ratios of the mutant:wt in the nasal washes of hamsters sampled 1–4 days after infection were measured by Sanger sequencing. Red dots represent individual animals ($n = 5$), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

Extended Data Fig. 4 Validation of N501Y competition assay by next generation sequencing.

a–e, The same RNA samples from N501Y competition with wt that were initially assessed using Sanger sequencing (Fig. 2a–f, Fig. 3g; **lower panels** here) were retested using next generation sequencing (NGS, **upper panels**). The samples were collected from the nasal washes (**a, b**) and tissues (**c, d**) from both donor and recipient hamsters and HAE cells (**e**). Red dots represent individual animals ($n = 5$, $n = 10$ in day 1 nasal wash of N501Y group) or biological repeats ($n = 6$ for HAE cells pooled from two independent experiments), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

Extended Data Fig. 5 Growth kinetics of mixed viruses in nasal wash of hamsters.

The replication kinetics of the mixed viruses (Alpha-FL with wt) in the nasal washes from both donors (**left panel**) and recipients (**right panel**) hamsters were measured by plaque assay. The nasal wash samples were collected from 1–4 days post-inoculation (donors) or post-contact (recipients). Dots represent individual hamsters ($n = 5$). The values in the graph represent the mean \pm standard error of the mean.

Extended Data Fig. 6 Competitions between Alpha-FL and Alpha-spike/WT-non-spike in hamsters.

a, b, The Alpha-FL was mixed with Alpha-spike/WT-non-spike virus and inoculated intranasally into hamsters. Results of the competition were assessed by sampling nasal washes of both donor (**a**) and recipient hamsters (**b**) from 1–4 days post-inoculation (donors) or post-contact (recipients). **c, d**, Results of the competition between the Alpha-FL and the Alpha-spike/WT-non-spike in the tracheae and lungs of both donor (**c**) and recipient hamsters (**d**) at 4 days post-inoculation or post-contact. **a–d**, Red dots represent individual animals ($n = 5$), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

Extended Data Fig. 7 The advantage of the N501Y substitution and Alpha-spike/WT-non-spike variant during the transmission from donor

to recipient hamsters.

a–d, The ratios of mixed viruses in the nasal washes (**a, c**), tracheae and lungs (**b, d**) of recipient hamsters were compared to the ratios of N501Y:wt or Alpha-spike/WT-non-spike:wt measured on the day 1 nasal wash of donor hamsters to assess fitness for transmission to and early replication in the recipient hamsters, respectively. The total infection titer of the mixed viruses was 10^5 PFU per hamster. Red dots represent individual animals ($n = 5$; $n = 10$ in day 1 nasal wash of N501Y group), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

Extended Data Fig. 8 Competition assay between Alpha-spike/WT-non-spike and WT on primary human airway epithelial cells.

The Alpha-spike/WT-non-spike and WT were mixed in 1:1 pfu ratio and inoculated onto human airway epithelial (HAE) cells at a total MOI of 5. The ratios of Alpha-spike/WT-non-spike to the wt virus were measured by Sanger sequencing. Red dots represent individual biological replicates ($n = 6$), pooled from 2 independent experiments. The horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

Extended Data Fig. 9 The spike N501Y substitution benefits viral infection of hamster upper airways.

a, Design of the hamster infection kinetic studies. The wt, N501Y and Alpha-spike viruses were intranasally inoculated into hamsters at a titer of 10^4 PFU per hamster. Nine hamsters were utilized for the initial infection in each group. At 2 days post-infection, 4 hamsters were sacrificed for organ collections. The nasal washes of the hamsters were collected on days 1, 2, 3 and 5 post-infection or before sacrifice. **b**, Weight change in hamsters following infection by the N501Y ($n = 5$) and Alpha-spike ($n = 5$) mutants compared to the wt ($n = 5$). MOCK group ($n = 4$) served as the negative (uninfected) control. The body weights were measured form 1–7 days post-infection. The weight loss data are shown as mean \pm standard deviation and statistically analyzed using two-factor analysis of variance (ANOVA) with Tukey's post hoc test. No significant differences were seen between the N501Y/Alpha-spike and wt groups. **c–h**, The infection of N501Y and Alpha-spike mutants compared to the wt in the nasal washes (**c–e**) collected 1 ($n = 9$), 2 ($n = 4$), 3 ($n = 5$), or 5 ($n = 5$) days

post-infection and in the organs (**f–h**) 2 days (n = 4) post-infection. The amounts of infectious virus (**c, f**) and genomic RNA (**d, g**) were quantified by plaque assay and RT–qPCR, respectively. The genomic RNA:PFU ratio (**e, h**) was calculated as an indication of virion infectivity. The values in the graph represent the mean ± standard error of the mean. A non-parametric two-tailed Mann-Whitney test was used to determine significant differences. *P* values were adjusted using the Bonferroni correction to account for multiple comparisons. Differences were considered significant if *p*<0.025.

Extended Data Table 1 The frequencies of the spike protein amino acid substitutions in SARS-CoV-2 strains

Supplementary information

[Supplementary Tables](#)

This file contains Supplementary Tables 1–3. Supplementary Table 1 contains input and output virus ratios in competition assays, Supplementary Table 2 shows total numbers of the 8 single substitutions in the Alpha spike gene among sequenced SARS-CoV-2 genomes and Supplementary Table 3 contains a list of primers used for mutation construction, RT–PCR, quantitative PCR and sequencing.

[Reporting Summary](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Liu, Y., Liu, J., Plante, K.S. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* **602**, 294–299 (2022).
<https://doi.org/10.1038/s41586-021-04245-0>

- Received: 09 March 2021
- Accepted: 12 November 2021
- Published: 24 November 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04245-0>

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

- [**Increased immune escape of the new SARS-CoV-2 variant of concern Omicron**](#)
 - Jie Hu
 - Pai Peng
 - Ai-long Huang

Cellular & Molecular Immunology (2022)
- [**Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta**](#)

- Lorenz Ulrich
- Nico Joel Halwe
- Charaf Benarafa

Nature (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04245-0>

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

- Article
- Open Access
- [Published: 25 November 2021](#)

Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation

- [Akatsuki Saito](#) ORCID: [orcid.org/0000-0002-2792-4399^{1,2,3 na1}](https://orcid.org/0000-0002-2792-4399),
- [Takashi Irie](#) ORCID: [orcid.org/0000-0003-3765-8482^{4 na1}](https://orcid.org/0000-0003-3765-8482),
- [Rigel Suzuki](#)^{5 na1},
- [Tadashi Maemura](#)^{6,7 na1},
- [Hesham Nasser](#) ORCID: [orcid.org/0000-0001-9163-1665^{8,9 na1}](https://orcid.org/0000-0001-9163-1665),
- [Keiya Uriu](#)^{10 na1},
- [Yusuke Kosugi](#)^{10 na1},
- [Kotaro Shirakawa](#) ORCID: [orcid.org/0000-0002-7469-1276¹¹](https://orcid.org/0000-0002-7469-1276),
- [Kenji Sadamasu](#) ORCID: [orcid.org/0000-0003-0541-706X¹²](https://orcid.org/0000-0003-0541-706X),
- [Izumi Kimura](#)¹⁰,
- [Jumpei Ito](#) ORCID: [orcid.org/0000-0003-0440-8321¹⁰](https://orcid.org/0000-0003-0440-8321),
- [Jiaqi Wu](#)^{13,14},
- [Kiyoko Iwatsuki-Horimoto](#) ORCID: [orcid.org/0000-0002-8266-020X⁶](https://orcid.org/0000-0002-8266-020X),
- [Mutsumi Ito](#)⁶,
- [Seiya Yamayoshi](#) ORCID: [orcid.org/0000-0001-7768-5157^{6,15}](https://orcid.org/0000-0001-7768-5157),
- [Samantha Loeber](#)¹⁶,
- [Masumi Tsuda](#) ORCID: [orcid.org/0000-0001-5400-5905^{17,18}](https://orcid.org/0000-0001-5400-5905),
- [Lei Wang](#) ORCID: [orcid.org/0000-0001-6145-395X^{17,18}](https://orcid.org/0000-0001-6145-395X),
- [Seiya Ozono](#) ORCID: [orcid.org/0000-0002-1764-2937¹⁹](https://orcid.org/0000-0002-1764-2937),
- [Erika P. Butlertanaka](#) ORCID: [orcid.org/0000-0002-7633-7238¹](https://orcid.org/0000-0002-7633-7238),
- [Yuri L. Tanaka](#) ORCID: [orcid.org/0000-0001-9866-4398¹](https://orcid.org/0000-0001-9866-4398),
- [Ryo Shimizu](#)^{8,20},
- [Kenta Shimizu](#) ORCID: [orcid.org/0000-0002-3825-7018⁵](https://orcid.org/0000-0002-3825-7018),
- [Kumiko Yoshimatsu](#) ORCID: [orcid.org/0000-0002-0062-2753²¹](https://orcid.org/0000-0002-0062-2753),
- [Ryoko Kawabata](#)⁴,
- [Takemasa Sakaguchi](#) ORCID: [orcid.org/0000-0002-3754-1252⁴](https://orcid.org/0000-0002-3754-1252),

- [Kenzo Tokunaga](#) ORCID: [orcid.org/0000-0002-2625-5322¹⁹](#),
- [Isao Yoshida¹²](#),
- [Hiroyuki Asakura](#) ORCID: [orcid.org/0000-0002-2572-6786¹²](#),
- [Mami Nagashima](#) ORCID: [orcid.org/0000-0001-8069-9250¹²](#),
- [Yasuhiro Kazuma](#) ORCID: [orcid.org/0000-0002-1046-8286¹¹](#),
- [Ryosuke Nomura¹¹](#),
- [Yoshihito Horisawa¹¹](#),
- [Kazuhisaya Yoshimura¹²](#),
- [Akifumi Takaori-Kondo](#) ORCID: [orcid.org/0000-0001-7678-4284¹¹](#),
- [Masaki Imai^{6,15}](#),
- [The Genotype to Phenotype Japan \(G2P-Japan\) Consortium](#),
- [Shinya Tanaka](#) ORCID: [orcid.org/0000-0001-6470-3301^{17,18}](#),
- [So Nakagawa](#) ORCID: [orcid.org/0000-0003-1760-3839^{13,14}](#),
- [Terumasa Ikeda](#) ORCID: [orcid.org/0000-0003-2869-9450⁸](#),
- [Takasuke Fukuhara](#) ORCID: [orcid.org/0000-0001-5471-8331⁵](#),
- [Yoshihiro Kawaoka](#) ORCID: [orcid.org/0000-0001-5061-8296^{6,7,15}](#) &
- [Kei Sato](#) ORCID: [orcid.org/0000-0003-4431-1380^{10,14}](#)

Nature volume 602, pages 300–306 (2022)

- 29k Accesses
- 14 Citations
- 625 Altmetric
- [Metrics details](#)

Subjects

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

Abstract

During the current coronavirus disease 2019 (COVID-19) pandemic, a variety of mutations have accumulated in the viral genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and, at the time of writing, four variants of concern are considered to be potentially hazardous to human society¹. The recently emerged B.1.617.2/Delta variant of concern is closely associated with the COVID-19 surge that occurred in India in the spring of 2021 (ref. ²). However, the virological properties of

B.1.617.2/Delta remain unclear. Here we show that the B.1.617.2/Delta variant is highly fusogenic and notably more pathogenic than prototypic SARS-CoV-2 in infected hamsters. The P681R mutation in the spike protein, which is highly conserved in this lineage, facilitates cleavage of the spike protein and enhances viral fusogenicity. Moreover, we demonstrate that the P681R-bearing virus exhibits higher pathogenicity compared with its parental virus. Our data suggest that the P681R mutation is a hallmark of the virological phenotype of the B.1.617.2/Delta variant and is associated with enhanced pathogenicity.

[Download PDF](#)

Main

During the current pandemic, SARS-CoV-2 has acquired a variety of mutations³. First, in spring 2020, a SARS-CoV-2 derivative containing a D614G mutation in its spike (S) protein emerged and quickly became predominant⁴. As the D614G mutation increases viral infectivity, fitness and interindividual transmissibility^{5,6,7,8,9,10}, the D614G-bearing variant quickly outcompeted the original strain. Since autumn 2020, some SARS-CoV-2 variants bearing multiple mutations have emerged and spread rapidly worldwide. As of September 2021, four variants of concern (VOCs) had emerged: B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma) and B.1.617.2 (Delta)^{11,12}.

The B.1.617 lineage emerged in India at the end of 2020 and is thought to have been a major driver of the massive COVID-19 surge in India that peaked at 400,000 infection cases per day². The B.1.617 lineage includes three sublineages—B.1.617.1, B.1.617.2 and B.1.617.3. Sublineage B.1.617.2 was defined as the latest VOC as of 25 November 2021, the Delta variant^{11,12}. Importantly, early evidence has suggested that infection with B.1.617.2/Delta may carry an increased risk of hospitalization compared with infection with B.1.1.7 (refs. ^{13,14,15}). However, the virological features of this newly emerging VOC, particularly its infectivity and pathogenicity, remain unclear. In this study, we demonstrate that B.1.617.2/Delta is more pathogenic than the prototypic SARS-CoV-2 in a Syrian hamster model. We also show that the P681R mutation in the S protein is a hallmark mutation of this lineage. The P681R mutation enhances the cleavage of the SARS-CoV-2 S protein and enhances viral fusogenicity. Moreover, we demonstrate that the P681R mutation can partly explain the higher pathogenicity of the B.1.617.2/Delta variant *in vivo*.

Epidemic dynamics of the B.1.617 lineage

We set out to investigate the phylogenetic relationships of the three subvariants belonging to the B.1.617 lineage. We downloaded 1,761,037 SARS-CoV-2 genomes

and corresponding data from the Global Initiative on Sharing All Influenza Data (GISAID) database (<https://www.gisaid.org>; as of 31 May 2021). As expected, each of the three sublineages B.1.617.1, B.1.617.2 and B.1.617.3 formed a monophyletic cluster (Fig. 1a and Extended Data Fig. 1). We next analysed the epidemic dynamics of each of the three B.1.617 sublineages. The B.1.617 variant, specifically B.1.617.1, was first detected in India on 1 December 2020 (GISAID ID: EPI_ISL_1372093) (Fig. 1b, c).

Fig. 1: Molecular phylogenetics and epidemic dynamics of the B.1.617 lineage pandemic.

 **figure 1**

a, Phylogenetic tree of the B.1.617 lineage. Scale bar, 0.0002 substitutions per site. Bootstrap values are indicated by asterisks; **100%, * $>70\%$. The uncollapsed tree is shown in Extended Data Fig. 1. **b, c**, Epidemic dynamics of the B.1.617 lineage. **b**, The number of sequences deposited in GISAID per day for India (top), the UK (middle) and the world (bottom). **c**, The percentages of each lineage deposited per day

from India (orange), the UK (blue) and the world (grey). The date on which each variant was first identified is indicated. The raw data are summarized in Supplementary Table 1. d, The proportion of amino acid replacements in the B.1.617 lineage. The top 10 replacements conserved in the S protein of B.1.617 and its sublineages are summarized. The numbers in parentheses indicate the number of sequences included in each panel. The raw data are summarized in Supplementary Table 2.

B.1.617.2 (GISAID ID: EPI_ISL_2131509) and B.1.617.3 (GISAID IDs: EPI_ISL_1703672, EPI_ISL_1703659 and EPI_ISL_1704392) were detected in India on 10 December 2020 and 13 February 2021, respectively (Fig. 1c). The prevalence of the B.1.617.1 sublineage peaked from February to April 2021 in India and then decreased (Fig. 1c). Although the B.1.617.3 variant has been detected sporadically in India, the B.1.617.2/Delta lineage has been dominant in India since March 2021 and has also spread all over the world (Fig. 1c). At the end of May 2021, 100%, 70% and 43.3% of the sequences deposited in GISAID per day from India (May 7), the UK (May 21) and the world (May 19), respectively, were B.1.617.2 sublineage sequences (Fig. 1c and Supplementary Table 1).

We next investigated the proportion of amino acid replacements in the S protein of each B.1.617 sublineage compared with the reference strain (Wuhan-Hu-1; GenBank: NC_045512.2). As shown in Fig. 1d, the L452R and P681R mutations were highly conserved in the B.1.617 lineage and, notably, the P681R mutation (16,650 out of 16,759 sequences, 99.3%) was the most representative mutation in this lineage. These data suggest that the P681R mutation is a hallmark of the B.1.617 lineage.

Syncytium formation by the Delta variant

To investigate the virological characteristics of the B.1.617.2/Delta variant, we conducted virological experiments using an isolate of B.1.617.2 (GISAID ID: EPI_ISL_2378732) as well as a D614G-bearing B.1.1 isolate (GISAID ID: EPI_ISL_479681) in Japan. In Vero cells, the growth of the B.1.617.2/Delta variant was significantly lower compared with the growth of the B.1.1 isolate (Fig. 2a). In particular, the viral RNA levels of the B.1.617.2/Delta variant at 48 h post-infection (h.p.i.) were more than 150-fold lower than those of the B.1.1 isolate (Fig. 2a). By contrast, although the growth kinetics of these viruses were relatively comparable in VeroE6/TMPRSS2 cells and Calu-3 cells (Fig. 2a), microscopy observations showed that the B.1.617.2/Delta variant formed larger syncytia than the B.1.1 virus (Fig. 2b). Measurements of the sizes of the floating syncytia in the infected VeroE6/TMPRSS2 culture indicated that the syncytia stimulated by B.1.617.2/Delta infection were significantly (3.6-fold) larger than those stimulated by B.1.1 infection (Fig. 2b). Moreover, the plaque size in VeroE6/TMPRSS2 cells infected with B.1.617.2/Delta

was significantly larger (1.2-fold) compared with in VeroE6/TMPRSS2 cells infected with B.1.1 virus (Extended Data Fig. [2a](#)). Immunofluorescence assays further showed that B.1.617.2/Delta-infected VeroE6/TMPRSS2 cells exhibited larger multinuclear syncytia compared with B.1.1-infected cells (Extended Data Fig. [3a](#)). Notably, although the B.1.1.7/Alpha and B.1.351/Beta VOCs also formed larger syncytia compared with B.1.1, the syncytia formed by B.1.617.2/Delta infection were 1.6-fold and 1.8-fold larger than those formed by B.1.1.7/Alpha and B.1.351/Beta infections, respectively, with statistical significance (Fig. [2b](#)). To directly assess the fusogenicity of the S proteins of these variants, we performed a cell-based fusion assay. We verified that this assay requires expression of human ACE2 in the target cells (Extended Data Fig. [4a](#)). Although the fusogenicity of S proteins of all VOCs tested was significantly greater than that of the parental D614G S, the B.1.617.2/Delta S exhibited the highest fusogenicity with statistical significance (Extended Data Fig. [4b](#)). These results suggest that the B.1.617.2/Delta variant promotes syncytium formation more strongly than the D614G-bearing B.1.1 virus as well as the B.1.1.7/Alpha and B.1.351/Beta VOCs.

Fig. 2: Virological features of the B.1.617.2/Delta variant in vitro and in vivo.

 [figure 2](#)

a, Growth kinetics of B.1.617.2/Delta variant. A B.1.617.2/Delta and a D614G-bearing B.1.1 were inoculated in cells, and the copy number of viral RNA in the supernatant was quantified using RT–qPCR. Assays were performed in quadruplicate. **b**,

Syncytium formation. Top, representative bright-field images of VeroE6/TMPRSS2 cells at 72 h.p.i. Scale bars, 100 µm. Bottom, the size distributions of floating syncytia in the cultures infected with B.1.1 ($n = 215$), B.1.1.7/Alpha ($n = 199$), B.1.351/Beta ($n = 249$) and B.1.617.2/Delta ($n = 216$). The size distribution of the floating uninfected cell culture ($n = 177$) is also shown as a negative control. **c–g**, Infection of Syrian hamsters with the B.1.617.2/Delta variant. Syrian hamsters were intranasally inoculated with B.1.1 ($n = 6$) and B.1.617.2/Delta ($n = 12$). Four hamsters of the same age were mock infected. The amount of viral RNA in the oral swab (**c**) and body weight (**d**) were measured. **e**, Haematoxylin and eosin (H&E) staining of the lungs of infected hamsters. Uninfected lung alveolar space and bronchioles are shown (left). Scale bars, 50 µm. **f**, Histopathological scoring of lung lesions. Representative pathological features are shown in Extended Data Fig. [5a](#). **g**, The area with large type II pneumocytes in the lungs of B.1.1-infected ($n = 4$) and B.1.617.2/Delta-infected ($n = 4$) hamsters at 5 d.p.i. The area was measured on the photographs (left) and summarized (right, each dot indicates the result from respective hamster). Raw data are shown in Extended Data Fig. [5b](#). Data are mean ± s.d. (**a, b**) or mean ± s.e.m. (**d, f, g**). In **a, b, g**, statistically significant differences versus B.1.1, B.1.1.7/Alpha and B.1.351/Beta (* $P < 0.05$) and uninfected culture (# $P < 0.05$) were determined using two-sided, unpaired Student's *t*-tests (**a, g**) or Mann–Whitney *U*-tests (**b**). In **c, d, f**, statistically significant differences between B.1.1 and B.1.617.2/Delta were determined by multiple regression and *P* values (**c, d**), and family-wise error rates calculated using the Holm method (**f**) are indicated in the figure. Statistically significant differences at each timepoint were also determined using two-sided unpaired Student's *t*-tests without adjustment for multiple comparisons (**c, d**), and those versus uninfected hamsters (* $P < 0.05$) are indicated by asterisks. The *P* value of the comparison between B.1.1 and B.1.617.2/Delta at each d.p.i. is indicated in the figure. NS, not significant.

Source data

The pathogenicity of the Delta variant

To investigate the pathogenicity of the B.1.617.2/Delta variant, we conducted hamster infection experiments using the B.1.617.2/Delta isolate and the B.1.1 isolate. The viral RNA loads in the oral swabs of B.1.617.2/Delta-infected hamsters were comparable with those of B.1.1-infected hamsters across timepoints on average ($P = 0.057$, multiple regression) (Fig. [2c](#)). Infected hamsters of both groups lost significant body weight beginning at 2 days post-infection (d.p.i.), and the weight loss of B.1.617.2/Delta-infected hamsters was significantly greater than that of B.1.1-infected hamsters across timepoints on average ($P = 0.0082$, multiple regression) (Fig. [2d](#)). The peak weight loss was 16% after infection with the B.1.617.2/Delta isolate and 13% for

the B.1.1 isolate; hamsters infected with the B.1.617.2/Delta isolate had a significantly greater weight loss compared with B.1.1 at 3 and 4 d.p.i. (Fig. 2d).

In the lungs of infected hamsters of both groups, bronchitis with focal inflammatory cell infiltration around bronchi/bronchioles was observed at 1 d.p.i. followed by haemorrhage or congestion at 3 d.p.i. (Fig. 2e,f and Extended Data Fig. 5a). Crushed nuclear debris, suggesting the damage of the alveolar pneumocytes with macrophage infiltration, was observed from 3 to 5 d.p.i., and the area of inflammatory cell infiltration was expanded with time (Fig. 2e,f and Extended Data Fig. 5a). In both cases, type II pneumocytes with an increased nuclear–cytoplasmic ratio appeared at 5 d.p.i. Notably, in the lungs of B.1.617.2/Delta-infected hamsters, prominently enlarged cells with large nuclei (greater than 10 µm in diameter) were recognized, suggesting large type II pneumocytes that reflect the severity of pneumonia at 5 d.p.i. (Fig. 2f,g and Extended Data Fig. 5b). Immunohistochemistry analysis of viral nucleocapsid (N) protein demonstrated that N proteins were detected in the bronchial epithelial cells with a small fraction of alveolar staining in both infection cases at 1 d.p.i. (Extended Data Fig. 5c). In the case of B.1.1 infection, N proteins were detected equally in bronchi/bronchioles at 1 and 3 d.p.i. (Extended Data Fig. 5c). At 5 d.p.i., alveolar pneumocytes exhibited positivity for N protein, which was weakened at 7 d.p.i. (Extended Data Fig. 5c). By contrast, in the case of B.1.617.2/Delta infection, the areas that were positive for N protein migrated rapidly to the alveolar pneumocytes around the bronchi/bronchioles and most of the bronchial epithelium was negative at 3 d.p.i. (Extended Data Fig. 5c). Thereafter, the N-positive areas further moved to the periphery of the lung lobes at 5 d.p.i. and were undetectable at 7 d.p.i. (Extended Data Fig. 5c). These observations suggest that the spatiotemporal distribution of infected cells between B.1.617.2/Delta and B.1.1 are different, and that the B.1.617.2/Delta isolate has higher pathogenicity in terms of the rapid spreading from bronchi/bronchioles to the alveolar space reaching the lung periphery than the B.1.1 isolate in spite of their relatively comparable proliferative potential.

The effect of the P681R mutation on viral fusion

The P681R mutation in the S protein is a unique feature of the B.1.617 lineage, including the B.1.617.2/Delta variant (Fig. 1d). As the P681R mutation is located in proximity to the furin cleavage site (FCS; residues RRAR positioned at 682–685) of the SARS-CoV-2 S protein¹⁶, we hypothesized that the P681R mutation is responsible for the promotion of cell–cell fusion, leading to the formation of larger syncytia. To address this possibility, we generated a P681R-bearing artificial virus by reverse genetics (Extended Data Fig. 2b) and performed further virological experiments. Although the amounts of viral RNA in the culture supernatants of the D614G/P681R-infected Vero and VeroE6/TMPRSS2 cells were significantly lower compared with those of the D614G-infected cells at some timepoints, the growth of these two viruses

was relatively comparable (Fig. [3a](#)). However, the floating syncytia (Fig. [3b](#)) and plaques (Extended Data Fig. [2c](#)) in the D614G/P681R-infected VeroE6/TMPRSS2 cells at 72 h.p.i. were significantly larger in size compared with the syncytia in the D614G-mutant-infected cells. Moreover, immunofluorescence assays showed that D614G/P681R-infected VeroE6/TMPRSS2 cells exhibited larger multinuclear cells than D614G-infected cells (Extended Data Fig. [3b](#)). These observations correspond well to the observations in the culture infected with the B.1.617.2/Delta variant (Fig. [2b](#) and Extended Data Figs. [2a](#), [3a](#)).

Fig. 3: Virological features of the P681R-containing virus in vitro.

 **figure 3**

a, The growth kinetics of artificially generated viruses. The D614G and D614G/P681R mutant viruses were generated by reverse genetics. These viruses (100 tissue culture infectious dose (TCID_{50})) were inoculated into Vero cells and VeroE6/TMPRSS2 cells, and the copy number of viral RNA in the culture supernatant

was quantified using RT–qPCR. The growth curves of the inoculated viruses are shown. Assays were performed in quadruplicate. **b**, **c**, Syncytium formation. **b**, Floating syncytia in VeroE6/TMPRSS2 cells infected with the D614G and D614G/P681R mutant viruses at 72 h.p.i. (top). Scale bars, 200 μ m. Bottom, the size distributions of floating syncytia in D614G-infected ($n = 228$) and D614G/P681R-infected ($n = 164$) cultures. **c**, Adherent syncytia in VeroE6/TMPRSS2 cells infected with GFP-expressing D614G- and D614G/P681R-mutant viruses at 24 h.p.i. Higher-magnification views of the regions indicated by with squares are shown in the right images. Scale bars, 200 μ m. The size distributions of adherent GFP⁺ syncytia in the D614G-infected ($n = 111$) and D614G/P681R-infected ($n = 126$) cultures. **d**, Western blot analysis of S-expressing cells. Left, representative blots of SARS-CoV-2 full-length S and cleaved S2 proteins as well as ACTB as an internal control. Assays were performed in triplicate. Data are mean \pm s.d. Right, the ratio of S2 to the full-length S plus S2 proteins in the S-expressing cells. **e**, SARS-CoV-2 S-based fusion assay. Effector cells (S-expressing cells) and target cells (ACE2-expressing cells or ACE2/TMPRSS2-expressing cells) were prepared, and the fusion activity was measured as described in the [Methods](#). Assays were performed in quadruplicate, and fusion activity (arbitrary units) is shown. Data are mean \pm s.d. Statistically significant differences versus D614G (* $P < 0.05$) and uninfected culture (# $P < 0.05$) were determined using two-sided unpaired Student's *t*-tests (**a**, **d**) or Mann–Whitney *U*-tests (**b**, **c**).

To clearly observe syncytium formation, we further generated GFP-expressing replication-competent D614G and D614G/P681R viruses. The levels of viral RNA in the supernatant and proportion of GFP-positive cells were similar in Vero, VeroE6/TMPRSS2 and Calu-3 cells (Extended Data Fig. [6](#)). However, at 24 h.p.i., significantly larger GFP-positive adherent syncytia were observed in VeroE6/TMPRSS2 cells infected with the GFP-expressing D614G/P681R virus (Fig. [3c](#)). Moreover, the GFP-positive floating syncytia at 72 h.p.i. in VeroE6/TMPRSS2 cells infected with GFP-expressing D614G/P681R virus were significantly larger (2.4-fold) in size compared with those of VeroE6/TMPRSS2 cells infected with GFP-expressing D614G virus (Extended Data Fig. [7a](#)). Moreover, GFP-positive syncytia were observed in D614G/P681R-infected Calu-3 cells but not in D614G-infected Calu-3 cells at 72 h.p.i. (Extended Data Fig. [6c](#)). These results suggest that the features of the B.1.617.2/Delta virus observed in in vitro cell culture experiments, particularly the formation of larger syncytia, are well reproduced by the insertion of the P681R mutation. To further investigate the effect of the P681R mutation, the GFP-expressing viruses were inoculated into human primary nasal epithelial culture. Notably, the viral RNA levels of D614G/P681R virus on the apical side of culture at 2 and 3 d.p.i. were 12.3-fold and 7.0-fold higher, respectively, than those of parental D614G virus with statistical significance, and the rapid growth of D614G/P681R virus was supported by the observation of GFP expression (Extended Data Fig. [7b](#)). Although the viral RNA

levels of D614G/P681R virus gradually decreased after 5 d.p.i., plaque-like spots were observed after 7 d.p.i., and the sizes of these plaque-like spots in the culture of D614G/P681R infection were significantly larger than the plaque-like spots in the culture of parental D614G virus infection (Extended Data Fig. 7b). These data suggest that the P681R mutation accelerates viral replication in human primary nasal epithelial culture and produces large plaque-like spots, which could be formed by cell-to-cell infection as the case of plaque formation.

To directly investigate the effect of the P681R mutation on the cleavage of the SARS-CoV-2 S protein, we prepared an HIV-1-based pseudovirus carrying the P681R mutation. Western blot analysis of the prepared pseudoviruses showed that the level of the cleaved S2 subunit was significantly increased in the presence of the P681R mutation (Extended Data Fig. 8a), suggesting that the P681R mutation facilitates furin-mediated cleavage of the SARS-CoV-2 S protein. We next performed a single-round pseudovirus infection assay using target HOS-ACE2 cells with or without *TMPRSS2* expression. The infectivity of both the D614G and D614G/P681R pseudoviruses was increased approximately tenfold by the expression of *TMPRSS2* in the target cells (Extended Data Fig. 8b). However, the relative infectivity of the D614G and D614G/P681R pseudoviruses was not altered by *TMPRSS2* expression (Extended Data Fig. 8b). These data suggest that the P681R mutation does not affect the infectivity of the viral particles.

We next addressed the effect of the P681R mutation on viral fusogenicity by a cell-based fusion assay. In the effector cells (that is, S-expressing cells), although the expression level of the D614G/P681R S protein was comparable to that of the D614G S protein, the level of the cleaved S2 subunit was significantly higher for the D614G/P681R mutant than for the D614G mutant (Fig. 3d). Consistent with the results of the pseudovirus assay (Extended Data Fig. 8a), these results suggest that the P681R mutation facilitates S cleavage. Flow cytometry analysis showed that the surface expression level of D614G/P681R S was significantly lower than that of D614G S (Extended Data Fig. 8c). Nevertheless, the cell-based fusion assay using the target cells without *TMPRSS2* demonstrated that D614G/P681R S is 2.1-fold more fusogenic than D614G S—a statistically significant difference ($P = 0.0002$, Welch's *t*-test) (Fig. 3e). Moreover, a mathematical modelling analysis of the fusion assay data showed that the initial fusion velocity of D614G/P681R S (0.83 ± 0.03 per hour) was significantly faster (2.8-fold) than that of D614G S (0.30 ± 0.03 per hour; $P = 4.0 \times 10^{-6}$, Welch's *t*-test) (Extended Data Fig. 8d, e). These data suggest that the P681R mutation enhances and accelerates SARS-CoV-2 S-mediated fusion. Furthermore, when we used targeted cells expressing *TMPRSS2*, both the fusion efficacy (about 1.2-fold) and initial fusion velocity (about 2.0-fold) were increased in both the D614G and D614G/P681R S proteins (Extended Data Fig. 8d, e). These results suggest that *TMPRSS2* facilitates the fusion mediated by SARS-CoV-2 S and human ACE2 and

that this TMPRSS2-dependent acceleration and promotion of viral fusion is not specific for the P681R mutant.

Neutralization of the P681R mutant

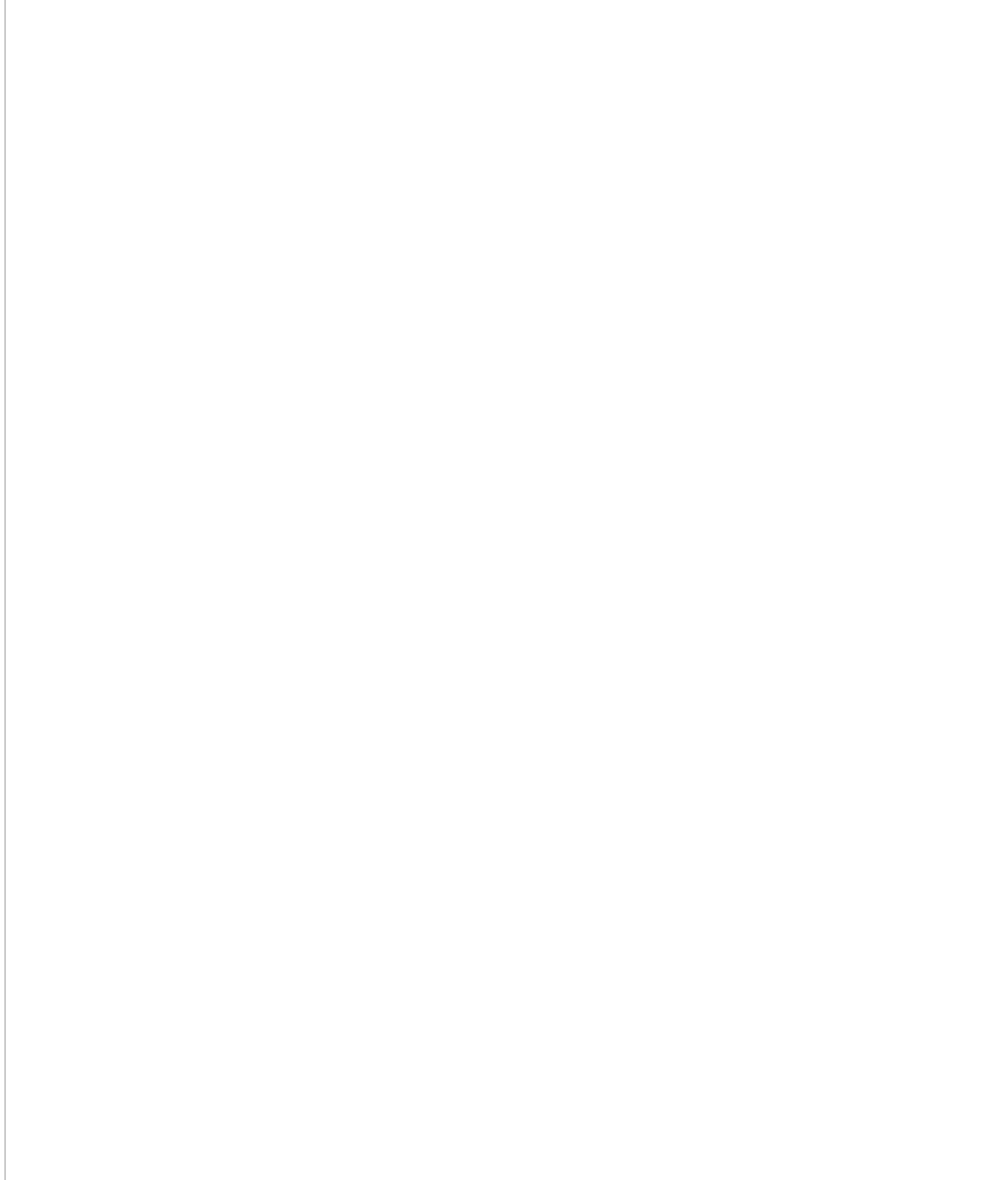
Resistance to neutralizing antibodies in the sera of COVID-19 convalescent individuals and vaccinated individuals is a hallmark of VOCs^{17,18}, and it has recently been shown that the B.1.617.2/Delta variant is relatively resistant to vaccine-induced neutralization^{19,20}. To determine whether the P681R mutation contributes to this virological phenotype, we performed a neutralization assay. The D614G/P681R pseudovirus was partially resistant (1.2–1.5-fold) to three monoclonal antibodies targeting the receptor-binding domain of the SARS-CoV-2 S protein (Extended Data Fig. 9a). Furthermore, neutralization experiments using 19 serum samples collected after two rounds of BNT162b2 vaccination showed that the D614G/P681R pseudovirus was significantly more resistant than the D614G pseudovirus to vaccine-induced neutralizing antibodies ($P < 0.0001$, Wilcoxon matched-pairs signed-rank test) (Extended Data Fig. 9b,c). These results suggest that the P681R-bearing pseudovirus is relatively resistant to neutralizing antibodies.

Pathogenicity of the P681R mutant

To assess the effect of the P681R mutation on viral replication and the pathogenicity of SARS-CoV-2, we intranasally infected Syrian hamsters with the D614G and D614G/P681R viruses. The D614G-infected hamsters exhibited no weight loss, although a slight decrease in body weight by 7 d.p.i. was observed for one of the hamsters (5.0%) (Fig. 4a). By contrast, all of the hamsters infected with the D614G/P681R virus experienced gradual body weight loss, and the hamsters showed a weight loss of 4.7–6.9% at 7 d.p.i., significantly greater compared with the weight loss of hamsters that were infected with the D614G virus ($P = 0.011$) (Fig. 4a). The weight loss of D614G/P681R-infected hamsters was significantly greater compared with that of D614G-infected hamsters on average across all timepoints ($P = 0.00015$, multiple regression) (Fig. 4a). We also assessed pulmonary function in infected hamsters by using a whole-body plethysmography system to measure enhanced pause (PenH), which is a surrogate marker of bronchoconstriction or airway obstruction. Infected hamsters of both groups showed increases in the lung PenH value, but the PenH values of D614G/P681R-infected hamsters were significantly higher than those of the D614G-infected hamsters on average across all timepoints ($P = 0.038$, multiple regression) (Fig. 4b). At 7 d.p.i., the D614G/P681R-infected hamsters had significantly higher PenH values than the D614G-infected hamsters ($P = 0.043$). At 3 d.p.i., both viruses replicated efficiently in the lungs and nasal turbinates of the infected hamsters, and no significant difference in viral replication was observed

between the two groups (Fig. 4c). At 7 d.p.i., no differences in viral titres in the nasal turbinates were found between the two groups; however, the lung titres in the D614G/P681R-infected group were significantly higher than those in the D614G-infected group ($P = 0.0013$) (Fig. 4c).

Fig. 4: Enhanced pathogenicity by the P681R mutation in hamsters.



Syrian hamsters were intranasally inoculated with the D614G and D614G/P681R viruses. **a**, Body weight changes in hamsters after viral infection. Body weights of virus-infected ($n = 4$ each) and uninfected ($n = 3$) hamsters were monitored daily for 7 days. **b**, Pulmonary function analysis in infected hamsters. Enhanced pause (PenH), which is a surrogate marker for bronchoconstriction or airway obstruction, was measured using whole-body plethysmography. **c**, Virus replication in infected hamsters. Four hamsters per group were euthanized at 3 d.p.i. and 7 d.p.i. for virus titration. Virus titres in the lungs (top) and nasal turbinates (bottom) were determined by plaque assay using VeroE6/TMPRSS2 cells. The points indicate data from individual Syrian hamsters. p.f.u., plaque-forming units. **d**, Histopathological examination of the lungs of infected Syrian hamsters. Representative pathological images of D614G- and D614G/P681R-infected lungs at 3 d.p.i. and 7 d.p.i. Scale bars, 200 μ m. Data are mean \pm s.e.m. In **a**, **b**, statistically significant differences were determined by multiple regression and P values are indicated in the figure. Statistically significant difference at each timepoint was also determined using two-sided unpaired Student's t -tests without adjustment for multiple comparisons, and those versus uninfected hamsters (* $P < 0.05$) are indicated by asterisks. The P value of the comparison between D614G and D614G/P681R at each d.p.i. is indicated in the figure.

[Source data](#)

Histopathological examination revealed cell infiltration in and around the bronchi/bronchioles at 3 d.p.i. in both groups, but solid bronchioloalveolar epithelial hyperplasia including type II pneumocytes was prominent at 7 d.p.i. in the D614G/P681R-infected hamsters (Fig. 4d). Microcomputed tomography (microCT) analysis revealed lung abnormalities in all of the infected hamsters on 7 d.p.i. that were consistent with commonly reported imaging features of COVID-19 pneumonia²¹ (Extended Data Fig. 10a). Lung abnormalities included multifocal nodular ground glass opacity with a peripheral, bilateral, multilobar, peribronchial distribution with regions of lung consolidation. The CT severity scores of the D614G-infected and D614G/P681R-infected hamsters ranged from 8 to 14, with an overall average CT severity score of 10.5 (median 9.5) (Extended Data Fig. 10b). The D614G/P681R-infected hamsters had a higher CT severity score (mean 11 (range 9–14, median 10.5)), compared with the D614G-infected hamsters (mean 10 (range 8–13, median 9.5)). Two of the D614G/P681R-infected hamsters developed a small-volume pneumomediastinum, probably secondary to severe pulmonary damage, micropulmonary rupture and gas tracking into the mediastinum.

Discussion

Previous studies have demonstrated the close association of FCS in the SARS-CoV-2 S protein with the viral replication mode and its dependence on TMPRSS2. Johnson et al.²³ and Peacock et al.²² showed that the loss of FCS results in an increase in viral replication efficacy in Vero cells and attenuates viral growth in Vero cells expressing TMPRSS2. By contrast, here we showed that the replication efficacy of the B.1.617.2/Delta variant was severely decreased in Vero cells compared with VeroE6/TMPRSS2 cells. Importantly, although FCS-deleted SARS-CoV-2 is less pathogenic compared with its parental virus²³, we revealed that the B.1.617.2/Delta variant and the P681R-harbouring virus exhibit higher pathogenicity. These findings suggest that enhanced viral fusogenicity, which is triggered by the P681R mutation, is closely associated with viral pathogenicity.

Although the P681R mutant is highly fusogenic, the virus containing the P681R mutation did not necessarily show stronger growth than the parental virus in in vitro cell cultures. HIV-1 variants with higher fusogenicity have been isolated from patients with AIDS, but the enhanced fusogenicity does not promote viral replication in in vitro cell cultures²⁴. Similarly, a measles virus (*Paramyxoviridae*) containing mutations in viral matrix proteins²⁵ and substitution mutations in viral fusion proteins^{26,27} is highly fusogenic and expands efficiently through cell–cell fusion. However, in in vitro cell cultures, the growth kinetics of these mutated measles viruses with higher fusogenicity are less efficient compared with those of the parental virus²⁵. Thus, the discrepancy between the efficacy of viral growth in in vitro cell cultures and viral fusogenicity is not unique to SARS-CoV-2. However, higher fusogenicity is associated with the severity of viral pathogenicity, such as in HIV-1 encephalitis²⁸ and the fatal subacute sclerosing panencephalitis that is caused by measles virus infection in the brain^{26,27}. Consistently, we showed that both the B.1.617.2/Delta variant and the P681R mutant exhibited higher fusogenicity in vitro and enhanced pathogenicity in vivo. Our data suggest that the greater COVID-19 severity and unusual symptoms caused by the B.1.617.2/Delta variant^{13,14,15} are due in part to the higher fusogenicity caused by the P681R mutation.

After launching this research in May 2021, the B.1.617.2/Delta variant has rapidly surpassed the other VOCs and is a major driver of the current COVID-19 pandemic worldwide in only a few months after the emergence. Revealing the rationale of higher transmissibility of this variant is one of the most urgent and crucial issues in the current COVID-19 pandemic. However, transmission experiments using the B.1.617.2/Delta variant or P681R-bearing virus in animal models were not performed in this study, and it remains unaddressed why the B.1.617.2/Delta variant has become more predominant than the other VOCs. By contrast, the greater severity and unusual COVID-19 symptoms caused by the B.1.617.2/Delta variant^{13,14,15} should be another important issue; we therefore addressed the virological properties and virulence of this variant and showed evidence suggesting that this pandemic variant has enhanced

fusogenicity and pathogenicity. We revealed the association of the P681R mutation with these virological features. An assumption from our observations is that the higher viral fusogenicity driven by the P681R mutation may be associated with the increased transmissibility of the B.1.617.2/Delta variant observed in humans. However, the P681R mutation is not specific for the B.1.617.2/Delta variant, and the sublineages related to the B.1.617.2/Delta variant, such as the B.1.617.1 and B.1.617.3 variants, that contain this mutation have not successfully spread in the human population. Thus, the mutations unique for the B.1.617.2/Delta variant would determine its higher transmissibility and further investigation will be needed to elucidate this property of the B.1.617.2/Delta variant.

Methods

Ethics statement

The virus isolation procedures in this study were approved by the Institutional Review Board of Tokyo Metropolitan Institute of Public Health according to the Declaration of Helsinki 2013 (3KenKenKen-466). All protocols involving specimens from human subjects recruited at Kyoto University were reviewed and approved by the Institutional Review Boards of Kyoto University (G0697) and the Institute of Medical Science, the University of Tokyo (2021-1-0416). All of the human participants provided written informed consent. All of the experiments with hamsters were performed in accordance with the Science Council of Japan's Guidelines for Proper Conduct of Animal Experiments. The protocols were approved by the Institutional Animal Care and Use Committee of National University Corporation Hokkaido University (20-0123) and the Animal Experiment Committee of the Institute of Medical Science, the University of Tokyo (PA19-75).

Collection of BNT162b2-vaccinated sera

Peripheral blood was collected four weeks after the second vaccination with BNT162b2 (Pfizer-BioNTech), and sera were isolated from the peripheral blood of 19 vaccinees (average age, 38; range, 28–59; 26% male). Sera were inactivated at 56 °C for 30 min and stored at –80 °C until use.

Cell culture

HEK293 cells (a human embryonic kidney cell line; ATCC CRL-1573), HEK293T cells (a human embryonic kidney cell line; ATCC CRL-3216) and HOS cells (a human osteosarcoma cell line; ATCC CRL-1543) were maintained in Dulbecco's modified Eagle's medium (high glucose) (Wako, 044-29765) containing 10% fetal bovine serum

(FBS) and 1% penicillin–streptomycin (PS). Vero cells (an African green monkey (*Chlorocebus sabaeus*) kidney cell line; JCRB0111) were maintained in Eagle’s minimum essential medium (Wako, 051-07615) containing 10% FBS and 1% PS. VeroE6/TMPRSS2 cells (an African green monkey(*C. sabaeus*) kidney cell line; JCRB1819)²⁹ were maintained in Dulbecco’s modified Eagle’s medium (low glucose) (Wako, 041-29775) containing 10% FBS, G418 (1 mg ml⁻¹; Nacalai Tesque, G8168-10ML) and 1% PS. Calu-3 cells (a human lung epithelial cell line; ATCC HTB-55) were maintained in Eagle’s minimum essential medium (Sigma-Aldrich, M4655-500ML) containing 10% FCS and 1% PS. HOS-ACE2/TMPRSS2 cells, HOS cells stably expressing human *ACE2* and *TMPRSS2*, were prepared as previously described^{30,31}. HEK293-C34 cells, *IFNARI*-KO HEK293 cells expressing human *ACE2* and *TMPRSS2* by doxycycline treatment³², were maintained in Dulbecco’s modified Eagle’s medium (high glucose) (Sigma-Aldrich, R8758-500ML) containing 10% FBS, 10 µg ml⁻¹ blasticidin (InvivoGen, ant-bl-1) and 1% PS. Primary human nasal epithelial cells (EP01, MD0436) were purchased from Epithelix and maintained according to the manufacturer’s procedure.

Animal experiments

Syrian hamsters (male, 4 weeks old) were purchased from Japan SLC. Baseline body weights were measured before infection. For the virus infection experiments in Fig. 2c, d, hamsters were euthanized by intramuscular injection of a mixture of 0.15 mg kg⁻¹ medetomidine hydrochloride (Domitor, Nippon Zenyaku Kogyo), 2.0 mg kg⁻¹ midazolam (Dormicum, Maruishi Pharmaceutical) and 2.5 mg kg⁻¹ butorphanol (Vetorphale, Meiji Seika Pharma). The B.1.1 or B.1.167.2/Delta viruses (10⁵ TCID₅₀ in 100 µl) were intranasally infected under anaesthesia. Body weights were measured, and oral swabs were collected under anaesthesia with isoflurane (Sumitomo Dainippon Pharma) daily. For the virus infection in Fig. 4, four hamsters per group were intranasally inoculated with the D614G or the D614G/P681R viruses (10⁴ TCID₅₀ in 30 µl) under isoflurane anaesthesia. Body weight was monitored daily for 7 days. For virological examinations, four hamsters per group were intranasally infected with the D614G or the D614G/P681R viruses (10⁴ TCID₅₀ in 30 µl); at 3 and 7 d.p.i., the hamsters were euthanized, and nasal turbinates and lungs were collected. The virus titres in the nasal turbinates and lungs were determined by plaque assays in VeroE6/TMPRSS2 cells.

Histopathological analysis

Excised animal tissues were fixed with 4% paraformaldehyde in PBS, and processed for paraffin embedding. The paraffin blocks were sectioned with a thickness of 3 µm and then mounted on silane-coated glass slides (MAS-GP, Matsunami). H&E staining

was performed according to a standard protocol. For immunohistochemistry analysis (Extended Data Fig. 5c), an Autostainer Link 48 (Dako) was used. The deparaffinized sections were exposed to EnVision FLEX target retrieval solution high pH (Agilent, K8004) for 20 min at 97 °C to activate, and a mouse anti-SARS-CoV-2 N monoclonal antibody (1:400 dilution; R & D systems, 1035111, MAB10474-SP) was used. The sections were sensitized using EnVision FLEX (Agilent) for 15 min and visualized by peroxidase-based enzymatic reaction with 3,3'-diaminobenzidine tetrahydrochloride as the substrate for 5 min.

Histopathological scoring of lung lesions

Pathological features, including bronchitis or bronchiolitis, haemorrhage or congestion, alveolar damage with epithelial apoptosis and macrophage infiltration, the presence of type II pneumocytes and the area of the presence of large type II pneumocytes (Fig. 2f and Extended Data Fig. 5b), were evaluated by certified pathologists and the degree of these pathological findings were arbitrarily scored using four-tiered system as 0 (negative), 1 (weak), 2 (moderate) and 3 (severe). Especially, for the evaluation of the area of the large type II pneumocytes at 5 d.p.i., the presence of more than 5 large type II pneumocytes with a nuclear diameter more than 10 µm per 0.04 mm² were delineated and the areas were measured using Fiji software v.2.2.0 implemented in ImageJ v.2.2.0.

Lung function

Respiratory parameters were measured using a whole-body plethysmography system (PrimeBioscience) according to the manufacturer's instructions. In brief, hamsters were placed in unrestrained plethysmography chambers and allowed to acclimatize for 1 min, then data were acquired over a 3 min period using FinePointe v.2.8.0.12146 (Data Sciences International).

MicroCT imaging

Respiratory organs of the infected hamsters were imaged by using an in vivo microCT scanner (CosmoScan GXII; Rigaku) at 7 d.p.i. Under ketamine–xylazine and isoflurane for the induction and maintenance of anaesthesia, the hamsters were placed in the imaging chamber and were scanned for 4 min at 90 kV, 88 µA, FOV 45 mm and a pixel size 90.0 µm. After scanning, the lung images were reconstructed and analysed using the CosmoScan Database software v.3.3.27.100 (Rigaku).

Qualitative and semiquantitative visual image analysis of the lungs was performed in three uninfected Syrian hamsters and the hamsters infected with D614G ($n = 4$) or D614G/P681R ($n = 4$) viruses at 7 d.p.i. A CT severity score (Extended Data Fig. 10b),

which was adapted from a human scoring system, was used to grade the severity of the lung abnormalities³³. Each lung lobe was analysed for degree of involvement and was scored from 0–4 as follows depending on the severity: 0 (none, 0%), 1 (minimal, 1%–25%), 2 (mild, 26%–50%), 3 (moderate, 51%–75%) or 4 (severe, 76%–100%). The scores for the five lung lobes were summed to obtain a total severity score of 0–20, reflecting the severity of abnormalities across the two infected groups. Images were anonymized and randomized; the scorer was blinded to the group allocation.

Viral genomes

All SARS-CoV-2 genome sequences and annotation information used in this study were downloaded from GISAID (<https://www.gisaid.org>) on 31 May 2021 (1,761,037 sequences). We first excluded genomes of viruses collected from non-human hosts. We obtained SARS-CoV-2 variants belonging to the B.1.617 lineage based on the PANGO annotation (that is, sublineages B.1.617.1, B.1.617.2/Delta or B.1.617.3) for each sequence in the GISAID metadata. One variant annotated as belonging to the B.1.617 lineage (GISAID ID: EPI_ISL_1544002, isolated in India on 25 February 2021) was not used in the analysis because the variant was not assigned to any of the three sublineages, possibly due to the 212 undetermined nucleotides in the genome. To infer the epidemiology of the B.1.617 lineage (Fig. 1b–d), we excluded genomes for which sampling date information was not available. We analysed 2,855, 13,821 and 83 sequences belonging to the B.1.617.1, B.1.617.2/Delta and B.1.617.3 sublineages, respectively.

A SARS-CoV-2 variant (GISAID ID: EPI_ISL_2220643) isolated in Texas, USA, on 10 August 2020, was also recorded to belong to B.1.617.1. However, the S protein of this viral sequence (GISAID ID: EPI_ISL_2220643) possesses neither L452R nor P681R mutations, both of which are features of the B.1.617 lineage. Thus, the EPI_ISL_2220643 sequence isolated in the USA is probably not the ancestor of the current B.1.617.1 lineage, and the EPI_ISL_1372093 sequence obtained in India can be considered to be the oldest example of the B.1.617 lineage.

Phylogenetic analyses

To infer the phylogeny of the B.1.617 sublineages, we screened SARS-CoV-2 genomes by removing genomes containing undetermined nucleotides at coding regions. As the numbers of genomes belonging to sublineages B.1.617.1 and B.1.617.2/Delta are large (894 and 6152 sequences, respectively), we used 150 randomly chosen sequences for each sublineage. For the B.1.617.3 sublineage, 32 genomes were used. We used the Wuhan-Hu-1 strain isolated in China on 31 December 2019 (GenBank: [NC_045512.2](#) and GISAID ID: EPI_ISL_402125) and the LOM-ASST-CDG1 strain isolated in Italy on 20 February 2020 (GISAID ID:

EPI_ISL_412973) together as an outgroup. We next collected 334 representative SARS-CoV-2 sequences and aligned the entire genome sequences using the FFT-NS-1 program in the MAFFT suite (v.7.407)³⁴. All sites with gaps in the alignment were removed, and the total length of the alignment was 29,085 nucleotides. A maximum likelihood tree was generated using IQ-TREE 2 v.2.1.3 with 1,000 bootstraps³⁵. The GTR+G substitution model was used based on the BIC criterion.

SARS-CoV-2 preparation and titration

A B.1.617.2/Delta isolate (strain TKYTK1734; GISAID ID: EPI_ISL_2378732) and a D614G-bearing B.1.1 isolate (strain TKYE610670; GISAID ID: EPI_ISL_479681) were isolated from SARS-CoV-2-positive individuals in Japan. In brief, 100 µl of nasopharyngeal swabs obtained from SARS-CoV-2-positive individuals were inoculated into VeroE6/TMPRSS2 cells in a biosafety level 3 laboratory. After incubation at 37 °C for 15 min, maintenance medium supplemented with Eagle's minimum essential medium (FUJIFILM Wako Pure Chemical, 056-08385) containing 2% FBS and 1% PS was added, and the cells were cultured at 37 °C under 5% CO₂. The cytopathic effect (CPE) was confirmed by observation under an inverted microscope (Nikon), and the viral load of the culture supernatant in which CPE was observed was confirmed by RT-qPCR. To determine the viral genome sequences, RNA was extracted from the culture supernatant using the QIAamp viral RNA mini kit (Qiagen, 52906). A cDNA library was prepared using the NEB Next Ultra RNA Library Prep Kit for Illumina (New England Biolab, E7530) and whole-genome sequencing was performed by MiSeq (Illumina).

To prepare the working virus stock, 100 µl of the seed virus was inoculated into VeroE6/TMPRSS2 cells (5×10^6 cells in a T-75 flask). Then, 1 h after infection, the culture medium was replaced with Dulbecco's modified Eagle's medium (low glucose) (Wako, 041-29775) containing 2% FBS and 1% PS. At 2–3 d.p.i., the culture medium was collected and centrifuged, and the supernatants were collected as the working virus stock.

The titre of the prepared working virus was measured as the 50% tissue culture infectious dose (TCID₅₀). In brief, 1 day before infection, VeroE6/TMPRSS2 cells (10,000 cells per well) were seeded into a 96-well plate. Serially diluted virus stocks were inoculated into the cells and incubated at 37 °C for 3 days. The cells were observed under microscopy to judge the CPE appearance. The value of TCID₅₀ per ml was calculated using the Reed–Muench method³⁶.

A B.1.1.7/Alpha isolate (strain QHN001; GISID ID: EPI_ISL_804007) and a B.1.351/Beta isolate (strain TY8-612; GISAID ID: EPI_ISL_1123289) were provided

by the National Institute for Infectious Diseases, Japan. The working viruses of these isolates were prepared as described above.

SARS-CoV-2 infection

One day before infection, Vero cells (10,000 cells), VeroE6/TMPRSS2 cells (10,000 cells) and Calu-3 cells (10,000 cells) were seeded into a 96-well plate. SARS-CoV-2 was inoculated and incubated at 37 °C for 1 h. The infected cells were washed, and 180 µl of culture medium was added. The culture supernatant (10 µl) was collected at the indicated timepoints, and RT–qPCR was used to quantify the viral RNA copy number (see below). To monitor the syncytium formation in infected cell culture, bright-field photos were obtained using ECLIPSE Ts2 (Nikon). The sizes of floating syncytia were measured using the ‘quick selection tool’ in Photoshop 2020 v.21.0.2 (Adobe) as pixels, and the areas of floating syncytia were calculated from the pixel value. For the GFP-expressing recombinant viruses (Extended Data Fig. [6c](#)), bright-field and green fluorescence images were obtained at the indicated timepoints using an All-in-One Fluorescence microscope BZ-X800 (Keyence), and the GFP fluorescence intensity was analysed using the BZ-X800 Analyzer v.1.1.2.4 (Keyence).

For the infection experiment primary human nasal epithelial cells (Extended Data Fig. [7b](#)), the working viruses were diluted with Opti-MEM (Thermo Fisher Scientific, 11058021). The diluted viruses (1,000 TCID₅₀ in 100 µl) were inoculated onto the apical side of the culture and incubated at 37 °C for 1 h. The inoculated viruses were removed and washed twice with Opti-MEM. To collect the viruses on the apical side of the culture, 100 µl Opti-MEM was applied onto the apical side of the culture and incubated at 37 °C for 10 min. Bright-field and green fluorescence images were obtained using ECLIPSE Ts2 (Nikon). The Opti-MEM applied was collected and we used RT–qPCR to quantify the viral RNA copy number (see below).

Immunofluorescence staining

One day before infection, VeroE6/TMPRSS2 cells (10,000 cells) were seeded into 96-well glass-bottom black plates and infected with SARS-CoV-2 (100 TCID₅₀). At 24 h.p.i., the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) (Nacalai Tesque, 09154-85) for 1 h at 4 °C. The fixed cells were permeabilized with 0.2% Triton X-100 in PBS for 1 h, blocked with 10% FBS in PBS for 1 h at 4 °C. The fixed cells were then stained using rabbit anti-SARS-CoV-2 N polyclonal antibody (GeneTex, GTX135570) for 1 h. After washing three times with PBS, cells were incubated with an Alexa 488-conjugated anti-rabbit IgG antibody (Thermo Fisher Scientific, A-11008) for 1 h. Nuclei were stained with DAPI (Thermo Fisher Scientific, 62248). Fluorescence microscopy was performed on an All-in-One Fluorescence Microscope BZ-X800 (Keyence).

Plaque assay

Plaque assay (Extended Data Fig. 2a, c) was performed as previously described³⁷. In brief, one day before infection, 200,000 VeroE6/TMPRSS2 cells were seeded into a 12-well plate. The virus was diluted with serum-free Dulbecco's modified Eagle's medium (low glucose) (Wako, 041-29775) containing 1% PS and 20 mM HEPES. After removing the culture medium, the cells were infected with 500 µl of the diluted virus at 37 °C. At 2 h.p.i., 1 ml of mounting solution containing 3% FCS and 1.5% carboxymethyl cellulose (Sigma-Aldrich, C9481-500G) was overlaid, followed by incubation at 37 °C. At 3 d.p.i., the culture medium was removed, and the cells were washed three times with PBS containing 0.9 mM calcium chloride and 0.5 mM magnesium chloride and fixed with 10% formaldehyde neutral buffer solution (Nacalai Tesque, 37152-51). The fixed cells were washed with tap water, dried and stained with staining solution (2% crystal violet (Nacalai Tesque, 09804-52) in water) for 30 min. The stained cells were washed with tap water and dried, and the size of plaques was measured using ImageJ.

SARS-CoV-2 reverse genetics

Recombinant SARS-CoV-2 was generated by circular polymerase extension reaction (CPER) as previously described^{32,37}. In brief, nine DNA fragments encoding the partial genome of SARS-CoV-2 (strain WK-521, PANGO lineage A; GISAID ID: EPI_ISL_408667)²⁹ were prepared by PCR using PrimeSTAR GXL DNA polymerase (Takara, R050A). A linker fragment encoding hepatitis delta virus ribozyme, bovine growth hormone polyA signal and cytomegalovirus promoter was also prepared by PCR. A summary of the corresponding SARS-CoV-2 genomic region and the PCR templates and primers used for this procedure is provided in Supplementary Table 3. The ten obtained DNA fragments were mixed and used for CPER³². To prepare GFP-expressing replication-competent recombinant SARS-CoV-2, we used fragment 9, in which the *GFP* gene was inserted into the *ORF7a* frame, instead of the authentic F9 fragment³² (Supplementary Table 3).

To produce recombinant SARS-CoV-2, the CPER products were transfected into HEK293-C34 cells using TransIT-LT1 (Takara, MIR2300) according to the manufacturer's protocol. At 1 day after transfection, the culture medium was replaced with Dulbecco's modified Eagle's medium (high glucose) (Sigma-Aldrich, R8758-500ML) containing 2% FCS, 1% PS and doxycycline (1 µg ml⁻¹; Takara, 1311N). At 6 days after transfection, the culture medium was collected and centrifuged, and the supernatants were collected as the seed virus. To remove the CPER products (that is, SARS-CoV-2-related DNA), 1 ml of the seed virus was treated with 2 µl TURBO DNase (Thermo Fisher Scientific, AM2238) and incubated at 37 °C for 1 h. Complete removal of the CPER products (that is, SARS-CoV-2-related DNA) from the seed

virus was verified by PCR. The working virus stock was prepared from the seed virus as described above.

To generate recombinant SARS-CoV-2 mutants, mutations were inserted in fragment 8 (Supplementary Table 3) using the GENEART site-directed mutagenesis system (Thermo Fisher Scientific, A13312) according to the manufacturer's protocol with the following primers: Fragment 8_S D614G forward, 5'-
CCAGGTTGCTGTTCTTATCAGGGTGTAACTGCACAGAACAGTCCCTG-3';
Fragment 8_S D614G reverse, 5'-
CAGGGACTTCTGTGCAGTTAACACCCTGATAAAGAACAGCAACCTGG-3';
Fragment 8_S P681R forward, 5'-
AGACTCAGACTAATTCTCGTCGGCGGGCACGTAGTGTA-3'; and Fragment 8_S P681R reverse, 5'-TACACTACGTGCCGCCGACGAGAATTAGTCTGAGTCT-3', according to the manufacturer's protocol. Nucleotide sequences were determined by a DNA sequencing service (Fasmac), and the sequencing data were analysed using Sequencher v.5.1 (Gene Codes). CPER for the preparation of SARS-CoV-2 mutants was performed using mutated fragment 8 instead of parental fragment 8. Subsequent experimental procedures were the same as those for parental SARS-CoV-2 preparation described above. To verify insertion of the mutation into the working viruses, viral RNA was extracted using the QIAamp viral RNA mini kit (Qiagen, 52906) and reverse-transcribed using SuperScript III reverse transcriptase (Thermo Fisher Scientific, 18080085) according to the manufacturers' protocols. DNA fragments including the inserted mutations were obtained by RT-PCR using PrimeSTAR GXL DNA polymerase (Takara, R050A) and the following primers: WK-521 23339-23364 forward, 5'-GGTGGTGTCACTGTTATAACACCAGG-3'; and WK-521 24089-24114 reverse, 5'-CAAATGAGGTCTCTAGCAGCAATATC-3'. Nucleotide sequences were confirmed as described above, and sequence chromatograms (Extended Data Fig. 2b) were visualized using the web application Tracy (<https://www.gear-genomics.com/teal/>)³⁸.

Viral genome sequencing analysis

The sequences of the working viruses were verified by viral RNA-sequencing analysis. Viral RNA was extracted using the QIAamp viral RNA mini kit (Qiagen, 52906). The sequencing library for total RNA sequencing was prepared using the NEB Next Ultra RNA Library Prep Kit for Illumina (New England Biolabs, E7530). Paired-end, 150 bp sequencing was performed using MiSeq (Illumina) with the MiSeq reagent kit v3 (Illumina, MS-102-3001). Sequencing reads were trimmed using fastp (v.0.21.0)³⁹ and subsequently mapped to the viral genome sequences of a lineage A isolate (strain WK-521; GISAID ID: EPI_ISL_408667)²⁹ or a GFP-inserted WK-521 (ref. 32) using BWA-MEM (v.0.7.17)⁴⁰. Variant calling, filtering and annotation were performed using SAMtools (v.1.9)⁴¹ and snpEff (v.5.0e)⁴². For the clinical isolates (a

B.1.617.2/Delta isolate (strain TKYTK1734; GISAID ID: EPI_ISL_2378732) and a D614G-bearing B.1.1 isolate (strain TKYE610670; GISAID ID: EPI_ISL_479681)), the detected variants that were present in the original sequences were excluded.

Information on the detected mutations is summarized in Supplementary Table 4, and the raw data are deposited at the Gene Expression Omnibus ([GSE182738](#)).

RT-qPCR

RT-qPCR was performed as previously described^{37,43}. In brief, 5 µl of culture supernatant was mixed with 5 µl of 2× RNA lysis buffer (2% Triton X-100, 50 mM KCl, 100 mM Tris-HCl (pH 7.4), 40% glycerol, 0.8 U µl⁻¹ recombinant RNase inhibitor (Takara, 2313B)] and incubated at room temperature for 10 min. RNase-free water (90 µl) was added, and the diluted sample (2.5 µl) was used as the template for RT-qPCR, which was performed according to the manufacturer's protocol using the One Step TB Green PrimeScript PLUS RT-PCR kit (Takara, RR096A) and the following primers: forward N, 5'-AGCCTCTTCGTTCCATCAC-3'; and Reverse N, 5'-CCGCCATTGCCAGCCATTC-3'. The viral RNA copy number was standardized using a SARS-CoV-2 direct detection RT-qPCR kit (Takara, RC300A). Fluorescent signals were acquired using the QuantStudio 3 Real-Time PCR system (Thermo Fisher Scientific), a CFX Connect Real-Time PCR Detection system (Bio-Rad), an Eco Real-Time PCR System (Illumina) or a 7500 Real Time PCR System (Applied Biosystems).

Plasmid construction

Plasmids expressing the SARS-CoV-2 S proteins of parental D614G (pC-SARS2-S D614G), B.1.1.7/Alpha (pC-SARS2-S Alpha), B.1.351/Beta (pC-SARS2-Beta) and B.1.617.2/Delta (pC-SARS2-S Delta) were prepared in a previous study^{31,44}. A plasmid expressing the SARS-CoV-2 S D614G/P681R mutant was generated by site-directed mutagenesis PCR using pC-SARS2-S D614G³¹ as the template and the following primers: P681R Fw, 5'-CCAGACCAACAGCCGGAGGAGGGCAAGGTCT-3' and P681R Rv, 5'-AGACCTTGCCTCCTCCGGCTGTTGGTCTGG-3'. The resulting PCR fragment was digested with KpnI and NotI and inserted into the KpnI-NotI site of the pCAGGS vector⁴⁵.

Pseudovirus assay

Pseudovirus assays were performed as previously described^{31,32,37}. In brief, lentivirus (HIV-1)-based luciferase-expressing reporter viruses pseudotyped with the SARS-CoV-2 S protein and its derivatives, HEK293T cells (1 × 10⁶ cells), were cotransfected

with 1 µg of psPAX2-IN/HiBiT⁴⁶, 1 µg of pWPI-Luc2⁴⁶ and 500 ng of plasmids expressing parental S or its derivatives using Lipofectamine 3000 (Thermo Fisher Scientific, L3000015) or PEI Max (Polysciences, 24765-1) according to the manufacturer's protocol. At 2 days after transfection, the culture supernatants were collected and centrifuged. The amount of pseudovirus prepared was quantified using the HiBiT assay as previously described^{31,46}. The prepared pseudoviruses were stored at -80 °C until use. For the experiment, HOS-ACE2 cells and HOS-ACE2/TMPRSS2 cells (10,000 cells per 50 µl) were seeded in 96-well plates and infected with 100 µl of pseudoviruses prepared at four different doses. At 2 d.p.i., the infected cells were lysed with a One-Glo luciferase assay system (Promega, E6130), and the luminescent signal was measured using a CentroXS3 plate reader (Berthold Technologies) or GloMax explorer multimode microplate reader 3500 (Promega).

Western blot analysis

Western blotting was performed as previously described^{47,48,49}. To quantify the level of the cleaved S2 protein in the cells, the collected cells were washed and lysed in lysis buffer (25 mM HEPES (pH 7.2), 20% glycerol, 125 mM NaCl, 1% Nonidet P40 substitute (Nacalai Tesque, 18558-54), protease inhibitor cocktail (Nacalai Tesque, 03969-21)). After quantification of total protein by protein assay dye (Bio-Rad, 50000006), lysates were diluted with 2× sample buffer (100 mM Tris-HCl (pH 6.8), 4% SDS, 12% β-mercaptoethanol, 20% glycerol, 0.05% bromophenol blue) and boiled for 10 min. Then, 10 µl samples (50 µg of total protein) were analysed using western blotting. To quantify the level of the cleaved S2 protein in the virions, 900 µl of the culture medium containing the pseudoviruses was layered onto 500 µl of 20% sucrose in PBS and centrifuged at 20,000g for 2 h at 4 °C. Pelleted virions were resuspended in 1× NuPAGE LDS sample buffer (Thermo Fisher Scientific, NP0007) containing 2% β-mercaptoethanol, and the lysed virions were analysed using western blotting. For protein detection, the following antibodies were used: mouse anti-SARS-CoV-2 S monoclonal antibody (1A9, GeneTex, GTx632604), rabbit anti-ACTB monoclonal antibody (13E5, Cell Signalling, 4970), mouse anti-HIV-1 p24 monoclonal antibody (183-H12-5C, obtained from the HIV Reagent Program, NIH, ARP-3537), horseradish peroxidase (HRP)-conjugated donkey anti-rabbit IgG polyclonal antibody (Jackson ImmunoResearch, 711-035-152) and HRP-conjugated donkey anti-mouse IgG polyclonal antibody (Jackson ImmunoResearch, 715-035-150). Chemiluminescence was detected using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, 34095) or Western BLoT Ultra Sensitive HRP Substrate (Takara, T7104A) according to the manufacturer's instructions. Bands were visualized using the Amersham Imager 600 (GE Healthcare), and the band intensity was quantified using Image Studio Lite v.5.2 (LI-COR Biosciences) or ImageJ v.2.2.0.

SARS-CoV-2 S-based fusion assay

The SARS-CoV-2 S-based fusion assay was performed as previously described³⁷. This assay uses a dual split protein (DSP) encoding *Renilla* luciferase (RL) and *GFP* genes; the respective split proteins, DSP_{1–7} and DSP_{8–11}, are expressed in effector and target cells by transfection^{48,50}. In brief, on day 1, effector cells (that is, S-expressing cells) and target cells (that is, ACE2-expressing cells) were prepared at a density of 0.6–0.8 × 10⁶ cells in a six-well plate. To prepare effector cells, HEK293 cells were cotransfected with 400 ng of the S expression plasmids and 400 ng pDSP_{1–7} using TransIT-LT1 (Takara, MIR2300). To prepare the target cells, HEK293 cells were cotransfected with pC-ACE2 (0 ng, 200 ng or 1,000 ng) and pDSP_{8–11} (400 ng). Target cells in selected wells were cotransfected with pC-TMPRSS2 (40 ng) in addition to the abovementioned plasmids. On day 3 (24 h after transfection), 16,000 effector cells were detached and reseeded into 96-well black plates (PerkinElmer, 6005225), and target cells were reseeded at a density of 1,000,000 cells per 2 ml per well in six-well plates. On day 4 (48 h after transfection), the target cells were incubated with EnduRen live cell substrate (Promega, E6481) for 3 h and then detached, and 32,000 target cells were added to a 96-well plate with effector cells. RL activity was measured at the indicated timepoints using a Centro XS3 LB960 (Berthold Technologies). The S proteins expressed on the surfaces of effector cells were stained with rabbit anti-SARS-CoV-2 S monoclonal antibody (HL6, GeneTex, GTX635654) or rabbit anti-SARS-CoV-2 S S1/S2 polyclonal antibody (Thermo Fisher Scientific, PA5-112048). Normal rabbit IgG (SouthernBiotech, 0111-01) was used as a negative control, and APC-conjugated goat anti-rabbit IgG polyclonal antibody (Jackson ImmunoResearch, 111-136-144) was used as a secondary antibody. Expression levels of surface S proteins were analysed using FACS Canto II (BD Biosciences) and FlowJo v.10.7.1 (BD Biosciences). RL activity was normalized to the mean fluorescence intensity of surface S proteins, and the normalized values are shown as fusion activity.

Mathematical modelling for fusion velocity quantification

The following cubic polynomial regression model was fitted to each of the time-series datasets (Fig. 3e):

$$y \approx b_0 + b_1 x + b_2 x^2 + b_3 x^3$$

The initial velocity of cell fusion was estimated from the derivative of the fitted cubic curve.

Neutralization assay

A virus neutralization assay was performed on HOS-ACE2/TMPRSS2 cells using SARS-CoV-2 S pseudoviruses expressing luciferase (see the ‘Pseudovirus assay’

section). The viral particles that were pseudotyped with D614G S or D614G/P681R S were incubated with serial dilutions of heat-inactivated human serum samples or three receptor-binding-domain-targeting neutralizing antibodies (8A5, Elabscience, E-AB-V1021; 4A3, Elabscience, E-AB-V1024; and CB6, Elabscience, E-AB-V1028) at 37 °C for 1 h. Pseudoviruses without sera and neutralizing antibodies were also included. An 80 µl mixture of pseudovirus and sera/neutralizing antibodies was then added to HOS-ACE2/TMPRSS2 cells (10,000 cells per 50 µl) in a 96-well white plate, and the luminescence was measured as described above (see the ‘Pseudovirus assay’ section). The 50% neutralization titre (NT₅₀) was calculated using Prism 9 software v.9.1.1 (GraphPad Software).

Statistics and reproducibility

In the time-course experiments using hamsters (Figs. 2c, d, f, 4a, b), two types of statistical tests were performed. First, to evaluate the difference between experimental conditions through all timepoints, a multiple regression analysis including experimental conditions as explanatory variables and timepoints as qualitative control variables was performed. *P* values were calculated using two-sided Wald tests. In Fig. 2f, family-wise error rates were calculated using the Holm method. Second, to evaluate the difference between two conditions at each timepoint, two-sided Student’s *t*-tests were performed. The data were analysed using Excel v.16.16.8 (Microsoft) or Prism 9 v.9.1.1 (GraphPad Software).

In Fig. 4d and Extended Data Fig. 5, the photographs shown are the representative areas of two independent experiments using 3 hamsters (6 lungs) at each timepoint. In Extended Data Fig. 3, assays were performed in quadruplicate. Photographs shown are the representative of 40 fields of view taken for each sample.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The raw data of virus sequences analysed in this study have been deposited at the Gene Expression Omnibus ([GSE182738](#)). Publicly available viral sequencing data are available from the GISAID database (<https://www.gisaid.org>). [Source data](#) are provided with this paper.

References

1. *Coronavirus Disease 2019* (WHO, 2021);
<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
2. Singh, J., Rahman, S. A., Ehtesham, N. Z., Hira, S. & Hasnain, S. E. SARS-CoV-2 variants of concern are emerging in India. *Nat. Med.* **27**, 1131–1133 (2021).
3. Plante, J. A. et al. The variant gambit: COVID's next move. *Cell Host Microbe* **29**, 508–515 (2021).
4. 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).
5. Zhou, B. et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **592**, 122–127 (2021).
6. Ozono, S. et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* **12**, 848 (2021).
7. Yurkovetskiy, L. et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* **183**, 739–751 (2020).
8. Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
9. Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2020).
10. Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75 (2021).
11. *SARS-CoV-2 Variant Classifications and Definitions*
<https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html> (CDC, 2021).
12. *Tracking SARS-CoV-2 Variants* <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (WHO, 2021).
13. *3 June 2021 Risk Assessment for SARS-CoV-2 Variant: Delta (VOC-21APR-02, B.1.617.2)*https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/991135/3%20June%202021%20Risk%20assessment%20for%20SARS-CoV-2%20variant%20DELTA.pdf (Public Health England, 2021).
14. Sheikh, A. et al. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* **397**, 2461–2462 (2021).

15. Twohig, K. A. et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 Delta (B.1.617.2) compared with Alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S0140-6736\(21\)01358-1](https://doi.org/10.1016/S0140-6736(21)01358-1) (2021).
16. Shang, J. et al. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl Acad. Sci. USA* **117**, 11727–11734 (2020).
17. Harvey, W. T. et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
18. Corti, D., Purcell, L. A., Snell, G. & Veesler, D. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* **184**, 3086–3108 (2021).
19. 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).
20. Planas, D. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021).
21. Simpson, S. et al. Radiological Society of North America expert consensus statement on reporting chest CT findings related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA—secondary publication. *J. Thorac. Imaging* **35**, 219–227 (2020).
22. Peacock, T. P. et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* **6**, 899–909 (2021).
23. Johnson, B. A. et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293–299 (2021).
24. Sterjovski, J. et al. Asn 362 in gp120 contributes to enhanced fusogenicity by CCR5-restricted HIV-1 envelope glycoprotein variants from patients with AIDS. *Retrovirology* **4**, 89 (2007).
25. Cathomen, T. et al. A matrix-less measles virus is infectious and elicits extensive cell fusion: consequences for propagation in the brain. *EMBO J.* **17**, 3899–3908 (1998).
26. Ikegami, S. et al. Fitness selection of hyperfusogenic measles virus F proteins associated with neuropathogenic phenotypes. *Proc. Natl Acad. Sci. USA* **118**, e2026027118 (2021).

27. Watanabe, S. et al. Mutant fusion proteins with enhanced fusion activity promote measles virus spread in human neuronal cells and brains of suckling hamsters. *J. Virol.* **87**, 2648–2659 (2013).
28. Rossi, F. et al. The V1-V3 region of a brain-derived HIV-1 envelope glycoprotein determines macrophage tropism, low CD4 dependence, increased fusogenicity and altered sensitivity to entry inhibitors. *Retrovirology* **5**, 89 (2008).
29. Matsuyama, S. et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl Acad. Sci. USA* **117**, 7001–7003 (2020).
30. Mlcochova, P. et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114–119 (2021).
31. Ozono, S. et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* **12**, 848 (2021).
32. Torii, S. et al. Establishment of a reverse genetics system for SARS-CoV-2 using circular polymerase extension reaction. *Cell Rep.* **35**, 109014 (2021).
33. Chung, M. et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* **295**, 202–207 (2020).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
36. Reed, L. J. & Muench, H. A simple method of estimating fifty percent endpoints. *Am. J. Hygiene* **27**, 493–497 (1938).
37. Motozono, C. et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **29**, 1124–1136 (2021).
38. Rausch, T., Fritz, M. H., Untergasser, A. & Benes, V. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. *BMC Genom.* **21**, 230 (2020).
39. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

40. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
41. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
42. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
43. Shema Mugisha, C. et al. A simplified quantitative real-time PCR assay for monitoring SARS-CoV-2 growth in cell culture. *mSphere* **5**, e00658-20 (2020).
44. Kimura, I. et al. SARS-CoV-2 Lambda variant exhibits higher infectivity and immune resistance. *Cell Rep.* <https://doi.org/10.1016/j.celrep.2021.110218> (2021).
45. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).
46. Ozono, S., Zhang, Y., Tobiume, M., Kishigami, S. & Tokunaga, K. Super-rapid quantitation of the production of HIV-1 harboring a luminescent peptide tag. *J. Biol. Chem.* **295**, 13023–13030 (2020).
47. Ikeda, T. et al. HIV-1 restriction by endogenous APOBEC3G in the myeloid cell line THP-1. *J. Gen. Virol.* **100**, 1140–1152 (2019).
48. Ikeda, T. et al. HIV-1 adaptation studies reveal a novel Env-mediated homeostasis mechanism for evading lethal hypermutation by APOBEC3G. *PLoS Pathog.* **14**, e1007010 (2018).
49. Sultana, T. et al. Multiple pathways to avoid beta interferon sensitivity of HIV-1 by mutations in capsid. *J. Virol.* **93**, e00986-19 (2019).
50. Kondo, N., Miyauchi, K. & Matsuda, Z. Monitoring viral-mediated membrane fusion using fluorescent reporter methods. *Curr. Protoc. Cell Biol.* **50**, 26.9.1–26.9.9 (2011).

Acknowledgements

We thank all of the members belonging to The Genotype to Phenotype Japan (G2P-Japan) Consortium; the staff at the National Institute for Infectious Diseases, Japan, and J. Gohda (The University of Tokyo, Japan) for providing virus isolates and

reagents; and all of the COVID-19 vaccine study participants at Kyoto University Hospital. An anti-HIV-1 p24 monoclonal antibody (183-H12-5C, ARP-3537) was obtained through the NIH HIV Reagent Program, NIAID, NIH (contributed by B. Chesebro and K. Wehrly). The supercomputing resource was provided by the Human Genome Center at The University of Tokyo and the NIG supercomputer at ROIS National Institute of Genetics. This study was supported in part by AMED Research Program on Emerging and Re-emerging Infectious Diseases (20fk0108163, to A.S.; 20fk0108401, to T.F.; 21fk0108617, to T.F.; 19fk0108113, to Y. Kawaoka; JP20fk0108412, to Y. Kawaoka; 20fk0108146, to K. Sato; 20fk0108270, to K. Sato; 20fk0108413, to T. Ikeda, S.N. and K. Sato; and 20fk0108451, to A.S., T. Irie, A.T.-K., G2P-Japan Consortium, S.N., T. Ikeda, T.F. and K. Sato); the AMED Research Program on HIV/AIDS (21fk0410033, to A.S.; and 21fk0410039, to K. Sato); the AMED Japan Program for Infectious Diseases Research and Infrastructure (20wm0325009, to A.S.; 21wm0325009, to A.S.; and 21wm0125002, to Y. Kawaoka); JST A-STEP (JPMJTM20SL, to T. Ikeda); JST SICORP (e-ASIA) (JPMJSC20U1, to K. Sato); JST SICORP (JPMJSC21U5, to K. Sato), JST CREST (JPMJCR20H6, to S.N.; and JPMJCR20H4, to K. Sato); JSPS KAKENHI Grant-in-Aid for Scientific Research C (19K06382, to A.S.; 18K07156, to K. Tokunaga; and 21K07060, to K. Tokunaga), Scientific Research B (18H02662, to K. Sato; and 21H02737, to K. Sato); the JSPS Fund for the Promotion of Joint International Research (Fostering Joint International Research) (18KK0447, to K. Sato); the JSPS Core-to-Core Program (JPJSCCA20190008, A. Advanced Research Networks, to K. Sato); the JSPS Research Fellow DC1 19J20488 (to I.K.); the JSPS Leading Initiative for Excellent Young Researchers (LEADER) (to T. Ikeda); the ONO Medical Research Foundation (to K. Sato); the Ichiro Kanehara Foundation (to K. Sato); the Lotte Foundation (to K. Sato); the Mochida Memorial Foundation for Medical and Pharmaceutical Research (to K. Sato); the Daiichi Sankyo Foundation of Life Science (to K. Sato); the Sumitomo Foundation (to K. Sato); the Uehara Foundation (to K. Sato); the Takeda Science Foundation (to T. Ikeda and K. Sato); The Tokyo Biochemical Research Foundation (to K. Sato); the Mitsubishi Foundation (to T. Ikeda); the Shin-Nihon Foundation of Advanced Medical Research (to T. Ikeda); the Tsuchiya Foundation (to T. Irie); a Grant for Joint Research Projects of the Research Institute for Microbial Diseases, Osaka University (to A.S.); an intramural grant from Kumamoto University COVID-19 Research Projects (AMABIE) (to T. Ikeda); the Intercontinental Research and Educational Platform Aiming for Eradication of HIV/AIDS (to T. Ikeda); and the Joint Usage/Research Center program of Institute for Frontier Life and Medical Sciences, Kyoto University (to K. Sato).

Author information

Author notes

1. These authors contributed equally: Akatsuki Saito, Takashi Irie, Rigel Suzuki, Tadashi Maemura, Hesham Nasser, Keiya Uriu, Yusuke Kosugi

Affiliations

1. Department of Veterinary Science, Faculty of Agriculture, University of Miyazaki, Miyazaki, Japan

Akatsuki Saito, Erika P. Butlertanaka, Yuri L. Tanaka & Nanami Morizako

2. Center for Animal Disease Control, University of Miyazaki, Miyazaki, Japan

Akatsuki Saito

3. Graduate School of Medicine and Veterinary Medicine, University of Miyazaki, Miyazaki, Japan

Akatsuki Saito

4. Institute of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

Takashi Irie, Ryoko Kawabata & Takemasa Sakaguchi

5. Department of Microbiology and Immunology, Graduate School of Medicine, Hokkaido University, Hokkaido, Japan

Rigel Suzuki, Kenta Shimizu, Haruko Kubo, Kana Tsushima & Takasuke Fukuhara

6. Division of Virology, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Tadashi Maemura, Kiyoko Iwatsuki-Horimoto, Mutsumi Ito, Seiya Yamayoshi, Masaki Imai & Yoshihiro Kawaoka

7. Influenza Research Institute, Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA

Tadashi Maemura & Yoshihiro Kawaoka

8. Division of Molecular Virology and Genetics, Joint Research Center for Human Retrovirus infection, Kumamoto University, Kumamoto, Japan

Hesham Nasser, Ryo Shimizu, Haruyo Hasebe, Kazuko Kitazato, Kohei Noda, Yue Yuan & Terumasa Ikeda

9. Department of Clinical Pathology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

Hesham Nasser

10. Division of Systems Virology, Department of Infectious Disease Control, International Research Center for Infectious Diseases, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Keiya Uriu, Yusuke Kosugi, Izumi Kimura, Jumpei Ito, Mika Chiba, Hirotake Furihata, Naoko Misawa, Akiko Oide, Mai Suganami, Miyabishara Yokoyama & Kei Sato

11. Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Kotaro Shirakawa, Yasuhiro Kazuma, Ryosuke Nomura, Yoshihito Horisawa & Akifumi Takaori-Kondo

12. Tokyo Metropolitan Institute of Public Health, Tokyo, Japan

Kenji Sadamasu, Isao Yoshida, Hiroyuki Asakura, Mami Nagashima & Kazuhisa Yoshimura

13. Department of Molecular Life Science, Tokai University School of Medicine, Kanagawa, Japan

Jiaqi Wu, Miyoko Takahashi & So Nakagawa

14. CREST, Japan Science and Technology Agency, Saitama, Japan

Jiaqi Wu, So Nakagawa & Kei Sato

15. The Research Center for Global Viral Diseases, National Center for Global Health and Medicine Research Institute, Tokyo, Japan

Seiya Yamayoshi, Masaki Imai & Yoshihiro Kawaoka

16. Department of Surgical Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, WI, USA

Samantha Loeber

17. Department of Cancer Pathology, Faculty of Medicine, Hokkaido University, Hokkaido, Japan

Masumi Tsuda, Lei Wang & Shinya Tanaka

18. Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Hokkaido, Japan

Masumi Tsuda, Lei Wang & Shinya Tanaka

19. Department of Pathology, National Institute of Infectious Diseases, Tokyo, Japan

Seiya Ozono & Kenzo Tokunaga

20. Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan

Ryo Shimizu

21. Institute for Genetic Medicine, Hokkaido University, Hokkaido, Japan

Kumiko Yoshimatsu

Consortia

The Genotype to Phenotype Japan (G2P-Japan) Consortium

- Mika Chiba
- , Hirotake Furihata
- , Haruyo Hasebe
- , Kazuko Kitazato
- , Haruko Kubo
- , Naoko Misawa
- , Nanami Morizako
- , Kohei Noda
- , Akiko Oide
- , Mai Suganami
- , Miyoko Takahashi
- , Kana Tsushima
- , Miyabishara Yokoyama
- & Yue Yuan

Contributions

A.S., T. Irie, R. Suzuki, H.N., K.U., Y. Kosugi, I.K., E.P.B., Y.L.T., R. Shimizu, K. Shimizu, R.K., T. Ikeda, T.F. and K. Sato performed cell culture experiments. R. Suzuki, T.M., K.I.-H., M. Ito, S.Y., K. Yoshimatsu, M. Imai, T.F. and Y. Kawaoka performed animal experiments. K. Sadamasu, S.O., T.S., K. Tokunaga, I.Y., H.A., M.N. and K. Yoshimura prepared experimental materials. K. Shirakawa, Y. Kazuma, R.N., Y.H. and A.T.-K. collected clinical samples. J.W. and S.N. performed molecular phylogenetic analysis. M. Tsuda, L.W. and S.T. performed histopathological analysis. S.L. performed microCT imaging analysis. J.I. performed mathematical and statistical analysis. A.S., T. Irie, M. Imai, S.T., S.N., T. Ikeda, T.F. Y. Kawaoka and K. Sato designed the experiments and interpreted the results. K. Sato wrote the original manuscript. All of the authors reviewed and proofread the manuscript. The Genotype to Phenotype Japan (G2P-Japan) Consortium contributed to the project administration.

Corresponding authors

Correspondence to [Shinya Tanaka](#), [So Nakagawa](#), [Terumasa Ikeda](#), [Takasuke Fukuhara](#), [Yoshihiro Kawaoka](#) or [Kei Sato](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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 A maximum-likelihood-based phylogenetic tree of 334 representative SARS-CoV-2 sequences belonging to the B.1.617](#)

lineage.

GISaid ID, country of exposure, and sampling date were noted in each terminal node. The country isolated (India, the UK, or other countries) and the PANGO sublineage are indicated by the text colour, as indicated in the figure. Coloured circles on the branch are shown on internal nodes for which the bootstrap value was ≥ 80 (red) or ≥ 50 (blue) ($n = 1,000$).

Extended Data Fig. 2 Plaques of SARS-CoV-2-infected VeroE6/TMPRSS2 cells.

a, A plaque assay was performed using VeroE6/TMPRSS2 cells as described in **Method**. Representative figures (top) and the summary of the size of plaques ($n = 20$ for each virus) are shown. Each dot indicates the diameter of the respective plaque. **b**, Chromatograms of nucleotide positions 23,399-23,407 (left) and 23,600-23,608 (right) of parental SARS-CoV-2 (strain WK-521, PANGO lineage A; GISaid ID: EPI_ISL_408667) and the D614G (A23403G in nucleotide) and P681R (C23604G in nucleotide) mutations. **c**, A plaque assay was performed using VeroE6/TMPRSS2 cells as described in **Method**. Representative figures (top) and the summary of the size of plaques ($n = 20$ for each virus) are shown. Each dot indicates the diameter of the respective plaque. Data are mean \pm S.D (**a, c**). Statistically significant differences between B.1.1 and B.1.617.2/Delta (**a**, $*P < 0.05$) and between D614G and D614G/P681R (**c**, $*P < 0.05$) were determined by two-sided Mann-Whitney U test.

Extended Data Fig. 3 Immunofluorescence staining of SARS-CoV-2-infected VeroE6/TMPRSS2 cells.

VeroE6/TMPRSS2 cells were infected with the B.1.1 or B.1.617.2/Delta (**a**) or artificially generated D614G or D614G/P681R (**b**) viruses [multiplicity of infection (MOI) 0.01]. The cells were stained with anti-SARS-CoV-2 nucleocapsid (N) (green) and DAPI (blue). Representative images taken at 24 h.p.i. Bars, 50 μ m.

Extended Data Fig. 4 SARS-CoV-2 S-based fusion assay.

a, Dependence of human ACE2 expression on the target cells for the SARS-CoV-2 S-based fusion assay. Target cells were prepared by transfecting the indicated amounts of human ACE2 expression plasmid, while Effector cells were prepared by transfecting SARS-CoV-2 S D614G expression plasmid. The fusion activity was measured as described in **Methods**. Assays were performed in quadruplicate, and fusion activity (arbitrary units) is shown. **b**, Fusogenic activity of the S proteins of VOCs. Effector cells (S-expressing cells) and target cells (ACE2-expressing cells) were prepared, and

the fusion activity was measured as described in [Methods](#). Note that the S protein sequence of “D614G” is identical to that of B.1.1 isolate. Assays were performed in quadruplicate, and fusion activity (arbitrary units) is shown. Data are mean \pm S.D. In **b**, statistically significant differences ($*P < 0.05$) versus the D614G (black), B.1.1.7/Alpha (blue) or B.1.351/Beta (green) were determined by two-sided Student’s *t* test.

[Extended Data Fig. 5 Histopathological features of lung lesions.](#)

a, Representative pathological features of lung including bronchitis/bronchiolitis, haemorrhage/congestion, alveolar damage with apoptosis and macrophage infiltration, presence of type II pneumocytes, and presence of the area of large type II pneumocytes are shown. 0 (negative), 1 (weak), 2 (moderate), and 3 (severe). Bars, 50 μm . **b**, Morphometrical analysis of the area of large type II pneumocytes. The area of the large type II pneumocytes with the nuclear diameter more than 8 μm in the lung specimens at 5 d.p.i. was measured, and the percentage of this area in the whole lung tissue area was calculated. Representative photographs of the lung tissue specimens with B.1.1 isolate (top) and B.1.617.2/Delta isolate (bottom) infections are shown. Red line indicates the area with the presence of large type II pneumocytes. Note that the most left panels (hamsters #89 and 90) are identical to the panels shown in Fig. [2g](#). **c**, IHC of the viral N proteins in the lung of infected hamsters. Representative IHC panels of the viral N proteins in the lung of hamsters infected with D614G-bearing B.1.1 isolate (left) and B.1.617.2/Delta isolates (right) are shown. Serial sections were used for H&E staining (top) and IHC (bottom). Bars, 250 μm (1, 3, and 7 d.p.i.) or 500 μm (5 d.p.i.).

[Extended Data Fig. 6 Growth kinetics of artificially generated GFP-expressing viruses.](#)

The GFP-expressing D614G and D614G/P681R mutant viruses were generated by reverse genetics. These viruses (100 TCID₅₀ for Vero and VeroE6/TMPRSS2 cells, 100 or 1,000 TCID₅₀ for Calu-3 cells) were inoculated into cells. The viral RNA copy number of in the culture supernatant (**a**) and the level of GFP-positive cells (the percentage of GFP-positive cells for Vero and VeroE6/TMPRSS2 cells; the GFP intensity per well for Calu-3 cells) (**b**) are shown. Data are mean \pm S.D. (**c**) Representative images of Calu-3 cells infected with GFP-expressing viruses (100 TCID₅₀). Areas enclosed with circles are enlarged in the right panels. Assays were performed in quadruplicate. Statistically significant differences ($*P < 0.05$) versus the D614G virus were determined by two-sided Student’s *t* test. NS, no statistical significance.

Extended Data Fig. 7 Syncytium formation in VeroE6/TMPRSS2 cells infected with GFP-expressing viruses.

a, (Left) Floating syncytia in VeroE6/TMPRSS2 cells infected with GFP-expressing D614G and D614G/P681R mutant viruses (100 TCID₅₀) at 72 h.p.i. Bars, 100 μm. (Right) The size distributions of adherent GFP⁺ syncytia in the D614G mutant-infected ($n = 147$) and the D614G/P681R mutant-infected ($n = 171$) cultures. **b**, The GFP-expressing D614G and D614G/P681R mutant viruses (1,000 TCID₅₀) were inoculated on the apical side of culture. (Upper left) The copy number of viral RNA on the apical side was quantified as described in [Methods](#), and the growth curves of the inoculated viruses are shown. (Lower left) The size distributions of plaque-like spots in D614G-infected and D614G/P681R-infected cultures. The numbers in the panel indicate the number of plaque-like spots counted. (Right) Time-course of GFP expression. Note that larger plaque-like spots are observed in D614G/P681R-infected culture after 7 d.p.i. Bars, 200 μm. Assays were performed in quadruplicate. Data are mean ± S.D. Statistically significant differences versus D614G (* $P < 0.05$) were determined by two-sided, unpaired Student's *t* test (**b**, upper left) or the Mann-Whitney U test (**a**, **b**, lower left).

Extended Data Fig. 8 Virological phenotypes exhibited by the P681R mutation.

a, Western blotting of pseudoviruses. (Left) Representative blots of SARS-CoV-2 full-length S and cleaved S2 proteins as well as HIV-1 p24 capsid as an internal control. kDa, kilodaltons. (Right) The ratio of S2 to the full-length S plus S2 proteins on pseudovirus particles. Assays were performed in triplicate. Data are mean ± S.D. A statistically significant difference (*, $P < 0.05$) versus D614G S was determined by two-sided Student's *t* test. **b**, Pseudovirus assay. The HIV-1-based reporter virus pseudotyped with SARS-CoV-2 S D614G or D614G/P681R was inoculated into HOS-ACE2 cells or HOS-ACE2/TMPRSS2 cells at 4 different doses (125, 250, 500 and 1,000 ng HIV-1 p24 antigen). Rates of infectivity compared to the virus pseudotyped with parental S D614G (1,000 ng HIV-1 p24) in HOS-ACE2 cells are shown. The labels above the HOS-ACE2/TMPRSS2 bars indicate the fold change versus the corresponding HOS-ACE2. Assays were performed in quadruplicate. **c**, Expression of S protein on the cell surface. (Left) Representative histogram of S protein expression on the cell surface. The number in the histogram indicates the mean fluorescence intensity (MFI). (Right) The MFI of surface S on the S-expressing cells. Assays were performed in triplicate. **d,e**, The kinetics of fusion velocity. **d**, Fitting of a mathematical model based on the kinetics of fusion activity data (see [Methods](#)). Each line indicates the result of respective mathematical model on the experimental data (shown in Fig. [3e](#)). **e**, Initial velocity of the S-mediated fusion. Assays were performed

in quadruplicate. Data are mean \pm S.D. Statistically significant differences ($*P < 0.05$) were determined by two-sided, unpaired Student's *t* test without adjustments for multiple comparisons (**b**), two-sided Student's *t* test (**c**) or two-sided Welch's *t* test (**e**). NS, no statistical significance.

Extended Data Fig. 9 Association of the P681R mutation with sensitivity to NAbs.

a, Neutralization assay using three monoclonal antibodies (clones 8A5, 4A3 and CB6). Assays were performed in triplicate. **b,c**, Neutralization assay using 19 vaccinated sera. Pseudoviruses and effector cells (S-expressing cells) were treated with serially diluted NAbS or sera as described in [Methods](#). Assays were performed in triplicate. The raw data of panel **b** are shown in panel **c**. NT₅₀, 50% neutralization titre. In **b**, each dot indicates the mean NT₅₀ value of the respective donor. A statistically significant difference versus the D614G virus was determined by two-sided Wilcoxon matched-pairs signed rank test. In **c**, the NT₅₀ values of D614G S (black) and D614G/P681R S (orange) for each serum are indicated in each panel.

Extended Data Fig. 10 MicroCT of the lung of infected hamsters.

a, MicroCT axial and coronal images of the lungs of Syrian hamsters at 7 d.p.i. with D614G-infected ($n = 4$), D614G/P681R-infected ($n = 4$), and uninfected hamsters ($n = 3$). Lung abnormalities included multifocal nodules (black arrows), ground glass opacity (white arrowheads), and regions of lung consolidation (white arrows) that were peripheral, bilateral, and multilobar. Pneumomediastinum is labelled with white asterisks. **b**, Summary of CT severity score. CT severity score of D614G-infected ($n = 4$), D614G/P681R-infected ($n = 4$), and uninfected hamsters ($n = 3$) Syrian hamsters. Each dot indicates the result of the respective infected hamster. Note that D614G/P681R-infected animals had a higher average CT severity score compared to D614G-infected animals.

Supplementary information

Supplementary Figures

Supplementary Fig. 1: gating strategy for flow cytometry of S protein expressing cells.
Supplementary Fig. 2: Original (uncrossed) blots.

Reporting Summary

Peer Review File

Supplementary Table 1

The number of daily deposited sequences in GISAID.

Supplementary Table 2

The percentage of mutations detected in the S protein of the B.1.617 lineage.

Supplementary Table 3

The SARS-CoV-2 genomic region encoded by each template and the primers used for the preparation of each fragment for CPER.

Supplementary Table 4

Summary of the viral sequences used in this study.

Source data

Source Data Fig. 2

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

Saito, A., Irie, T., Suzuki, R. *et al.* Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* **602**, 300–306 (2022).
<https://doi.org/10.1038/s41586-021-04266-9>

- Received: 22 July 2021
- Accepted: 18 November 2021
- Published: 25 November 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04266-9>

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

Further reading

- [**Peptide-based pan-CoV fusion inhibitors maintain high potency against SARS-CoV-2 Omicron variant**](#)
 - Shuai Xia
 - Jasper Fuk-Woo Chan
 - Shibo Jiang

Cell Research (2022)

- **Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19**

- Can-rong Wu
- Wan-chao Yin
- H. Eric Xu

Acta Pharmacologica Sinica (2022)

- **Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2**

- Kapil Gupta
- Christine Toelzer
- Imre Berger

Nature Communications (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04266-9>

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

- Article
- Open Access
- [Published: 22 December 2021](#)

Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta

- [Lorenz Ulrich](#) [ORCID: orcid.org/0000-0001-5004-806X](#)¹ nal,
- [Nico Joel Halwe](#) [ORCID: orcid.org/0000-0002-7983-2808](#)¹ nal,
- [Adriano Taddeo](#) [ORCID: orcid.org/0000-0003-3281-1548](#)^{2,3} nal,
- [Nadine Ebert](#)^{2,3} nal,
- [Jacob Schön](#)¹,
- [Christelle Devisme](#)^{2,3},
- [Bettina Salome Trüeb](#)^{2,3,4},
- [Bernd Hoffmann](#) [ORCID: orcid.org/0000-0001-5358-6445](#)¹,
- [Manon Wider](#)⁵,
- [Xiaoyu Fan](#)⁶,
- [Meriem Bekliz](#)⁷,
- [Manel Essaidi-Laziosi](#)⁷,
- [Marie Luisa Schmidt](#)⁸,
- [Daniela Niemeyer](#) [ORCID: orcid.org/0000-0002-1897-6365](#)^{8,9},
- [Victor Max Corman](#) [ORCID: orcid.org/0000-0002-3605-0136](#)^{8,9},
- [Anna Kraft](#)¹,
- [Aurélie Godel](#)^{2,3},
- [Laura Laloli](#)^{5,10},
- [Jenna N. Kelly](#) [ORCID: orcid.org/0000-0001-7265-2233](#)^{2,3},
- [Brenda M. Calderon](#)⁶,
- [Angele Breithaupt](#) [ORCID: orcid.org/0000-0002-6373-5923](#)¹¹,
- [Claudia Wylezich](#) [ORCID: orcid.org/0000-0002-0436-4480](#)¹,
- [Inês Berenguer Veiga](#) [ORCID: orcid.org/0000-0001-8554-790X](#)^{2,3},
- [Mitra Gultom](#)^{5,10},
- [Sarah Osman](#)⁶,
- [Bin Zhou](#) [ORCID: orcid.org/0000-0002-1535-6283](#)⁶,
- [Kenneth Adeea](#)⁷,
- [Benjamin Meyer](#) [ORCID: orcid.org/0000-0003-0601-3550](#)¹²,

- [Christiane S. Eberhardt](#)^{12,13,14},
- [Lisa Thomann](#) ORCID: orcid.org/0000-0003-3530-0524^{2,3},
- [Monika Gsell](#)⁵,
- [Fabien Labroussaa](#) ORCID: orcid.org/0000-0002-8712-3024⁴,
- [Jörg Jores](#) ORCID: orcid.org/0000-0003-3790-5746⁴,
- [Artur Summerfield](#)^{2,3},
- [Christian Drosten](#) ORCID: orcid.org/0000-0001-7923-0519^{8,9},
- [Isabella Anne Eckerle](#)^{7,15,16},
- [David E. Wentworth](#) ORCID: orcid.org/0000-0002-5190-980X⁶,
- [Ronald Dijkman](#) ORCID: orcid.org/0000-0003-0320-2743^{5 na2},
- [Donata Hoffmann](#) ORCID: orcid.org/0000-0003-4552-031X^{1 na2},
- [Volker Thiel](#) ORCID: orcid.org/0000-0002-5783-0887^{2,3 na2},
- [Martin Beer](#) ORCID: orcid.org/0000-0002-0598-5254^{1 na2} &
- [Charaf Benarafa](#) ORCID: orcid.org/0000-0002-2049-7769^{2,3 na2}

Nature volume 602, pages 307–313 (2022)

- 8887 Accesses
- 110 Altmetric
- [Metrics details](#)

Subjects

- [Experimental models of disease](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

Abstract

Emerging variants of concern (VOCs) are driving the COVID-19 pandemic^{1,2}. Experimental assessments of replication and transmission of major VOCs and progenitors are needed to understand the mechanisms of replication and transmission of VOCs³. Here we show that the spike protein (S) from Alpha (also known as B.1.1.7) and Beta (B.1.351) VOCs had a greater affinity towards the human angiotensin-converting enzyme 2 (ACE2) receptor than that of the progenitor variant S(D614G) *in vitro*. Progenitor variant virus expressing S(D614G) (wt-S^{614G}) and the Alpha variant showed similar replication kinetics in human nasal airway epithelial cultures, whereas the Beta variant was outcompeted by both. *In vivo*, competition experiments showed a clear fitness advantage of Alpha over wt-S^{614G} in ferrets and

two mouse models—the substitutions in S were major drivers of the fitness advantage. In hamsters, which support high viral replication levels, Alpha and wt-S^{614G} showed similar fitness. By contrast, Beta was outcompeted by Alpha and wt-S^{614G} in hamsters and in mice expressing human ACE2. Our study highlights the importance of using multiple models to characterize fitness of VOCs and demonstrates that Alpha is adapted for replication in the upper respiratory tract and shows enhanced transmission in vivo in restrictive models, whereas Beta does not overcome Alpha or wt-S^{614G} in naive animals.

[Download PDF](#)

Main

Uncontrolled transmission of SARS-CoV-2 in the human population has contributed to the persistence of the COVID-19 pandemic. The emergence of new variants in largely immunologically naive populations suggests that adaptive mutations in the viral genome continue to improve the fitness of this zoonotic virus. In March 2020, a single amino acid change in the S protein at position 614 (S(D614G)) was identified in a small fraction of sequenced samples—this became the predominant variant worldwide within a few weeks⁴. The fitness advantage conferred by this single amino acid change was supported by major increases in infectivity, viral load and transmissibility in vitro and in animal models^{3,5,6}.

In the second half of 2020, SARS-CoV-2 VOCs with a combination of several mutations emerged, including Alpha, first described in southeast England⁷, and Beta, first identified in South Africa⁸. In February–March 2021, Alpha rapidly became the prevailing variant in many regions of the world and a higher reproduction number was inferred from early epidemiological data^{9,10,11}. Beyond S(D614G), Alpha has 18 further mutations in its genome compared with the progenitor, with two deletions and six substitutions within the S gene¹². Some of the S mutations, such as N501Y and the H69/V70 deletion, have been hypothesized to enhance replication and transmission, but there is a lack of clear experimental evidence for this^{13,14}. Beta has nine mutations in S, including N501Y, and two in the S receptor-binding domain (RBD), K417N and E484K. E484K is thought to be responsible for the ability of Beta to escape neutralization by plasma from convalescent individuals^{15,16,17}. Whether S mutations are solely responsible for the putative fitness advantage and if so, which ones, remains unknown.

Here we investigate the fitness of Alpha and Beta VOCs relative to wt-S^{614G}, the predominant parental strain containing the S(D614G) substitution—in relevant primary airway culture systems in vitro, and in ferrets, Syrian hamsters and two mouse

models expressing human ACE2—to assess specific advantages in replication and transmission and to evaluate the effects of Alpha S mutations alone in vivo. Neither Alpha nor Beta showed enhanced replication in human airway epithelial cell (AEC) cultures compared with wt-S^{614G}. Competitive transmission experiments in Syrian hamsters showed similar replication and transmission of wt-S^{614G} and Alpha, which both outcompeted Beta. However, competitive experiments in ferrets and transgenic mice expressing human ACE2 controlled by the *KRT18* promoter (hACE2-K18Tg), which overexpress human ACE2 in epithelial cells, showed increased fitness of Alpha compared with wt-S^{614G}. Finally, Alpha and a recombinant clone of progenitor virus expressing the Alpha S protein (wt-S^{Alpha}) both outcompeted the parental wt-S^{614G} strain, resulting in higher virus load in the upper respiratory tract (URT) of mice expressing human ACE2 instead of mouse ACE2 under the endogenous mouse *Ace2* promoter (hACE2-KI mice). Similar to results from AEC cultures, Beta showed lower fitness than wt-S^{614G} in hACE2-KI mice. Infections with Alpha and wt-S^{614G} virus resulted in similar pathologies in all the in vivo models.

Binding and replication of VOCs in vitro

The evolution of SARS-CoV-2 variants is associated with accumulation of mutations in the S protein. We determined dissociation constant (K_D) values for recombinant trimeric S with immobilized dimeric human ACE2 using bio-layer interferometry. S protein from Alpha (S^{Alpha}) or Beta (S^{Beta}) exhibited a fourfold higher affinity for human ACE2 than that of S(D614G) protein (Extended Data Fig. 1a). Replication kinetics of Alpha, Beta and a wild-type clinical isolate with the S(D614G) mutation were similar in relation to viral copies and titres in AEC cultures incubated at 33 and 37 °C (Extended Data Fig. 1b). However, in direct competition experiments in AEC cultures, Alpha had no advantage over wt-S^{614G}, whereas Beta was outcompeted by both Alpha and wt-S^{614G} (Extended Data Fig. 1c), indicating that competition experiments can expose differences in replication that are not detected in individual growth kinetic assays.

Alpha and wt-S^{614G} outcompete Beta in hamsters

We inoculated groups of six Syrian hamsters intranasally with a mixture of two SARS-CoV-2 strains comprising equivalent numbers of genome copies in three one-to-one competition experiments: Alpha versus Beta, Beta versus wt-S^{614G}, and Alpha versus wt-S^{614G}. All experimentally infected ‘donor’ hamsters were kept strictly in isolation cages to prevent intergroup spill-over infections. Each donor hamster was co-housed with a naive ‘contact 1’ hamster 1 day post infection (dpi), creating six donor–contact 1 pairs to evaluate shedding and transmission. At 4 dpi, donor hamsters were

euthanized and six subsequent transmission pairs were set up by co-housing each contact 1 hamster with a naive contact 2 hamster (Extended Data Fig. [2a](#)).

In two competition experiments, wt-S^{614G} and Alpha outcompeted Beta, as indicated by nasal washes of the donor hamsters from 1 dpi until euthanasia at 4 dpi. The viral load reached 10⁹ genome copies (gc) per ml for wt-S^{614G} and Alpha, whereas Beta viral loads were tenfold lower at corresponding time points. Consequently, transmission of Beta was limited or undetectable in contact 1 and contact 2 hamsters compared with the competing variants wt-S^{614G} (Fig. [1](#)) and Alpha (Extended Data Fig. [3](#)). Transmission to contact hamsters was associated with clinical signs and weight loss (Extended Data Fig. [4a,b](#)). In donor and contact hamsters, viral genome loads in the URT (comprising nasal conchae and trachea) revealed increased replication of Alpha and wt-S^{614G} compared with Beta (Extended Data Fig. [5a,b](#)), which may explain the lower transmission rate of Beta in a competition context. Of note, Beta replicated to high titres in the lower respiratory tract (LRT; comprising cranial, medial and caudal lung lobes) of donor hamsters, similar levels as observed for the competing Alpha and wt-S^{614G} virus (Extended Data Fig. [5a,b](#)).

Fig. 1: Competitive replication and transmission of Beta and wt-S^{614G} in Syrian hamsters.

 **figure 1**

Six donor hamsters were each inoculated with a median tissue culture infectious dose (TCID₅₀) of 10^{4.25}, determined by back titration and comprising a mixture of wt-S^{614G} (orange) and Beta (green) at a 1:3.8 ratio, determined by quantitative PCR with reverse transcription (RT–qPCR). Donor, contact 1 and contact 2 hamsters were co-housed sequentially as shown in Extended Data Fig. [2a](#). Nasal washes were performed daily from 1–9 dpi and then every 2 days until 21 dpi. Pie charts show the ratio of variants detected in nasal washes at the indicated dpi. Pie chart sizes are proportional to the total number of viral genome copies per ml, as shown above or below each chart. Grey pies indicate values below the limit of detection (LOD; <10³ viral genome copies per ml). Hamster silhouettes are coloured according to the dominant variant (>66%) detected in the last positive sample from each animal. Daggers indicate that the animal reached the humane endpoint; double daggers indicate a hamster that died during inhalation anaesthesia at 3 and 4 dpi. This required changes in the group

composition in cage 6—the donor hamster was kept until 7 dpi and was co-housed in two different pairs: donor–contact 1a and donor–contact 1b.

Source data

Competition between Alpha and wt-S^{614G} showed no clear difference in virus titres in nasal washes of donor hamsters, and both variants were detected at all time points in each donor with numbers of individual variants ranging from 10⁵ to 10⁹ gc ml⁻¹ (Fig. 2). Of note, Alpha was dominant over wt-S^{614G} in the donor hamsters at 1 dpi, but these strains were balanced by the endpoint at 4 dpi. In organ samples from the donor hamsters, the highest viral loads were found in the LRT, where Alpha was predominant (more than 66% of genome copies) overall with more than tenfold more viral genome copies than wt-S^{614G} in 14 out of 18 lung samples from the 6 donor hamsters (Extended Data Fig. 5c). Sequential transmission to contact animals was associated with body weight loss (Extended Data Fig. 4c) and was highly efficient for Alpha and wt-S^{614G} variants, which were both detected in nasal washes of almost all contact 1 hamsters (Fig. 2). Whereas all donor and contact 1 hamsters transmitted both viruses to their respective contacts, contact 2 hamsters mainly shed one variant at high levels in nasal washes, demonstrating similar transmission ability for wt-S^{614G} and Alpha. At the individual endpoints for contact 1 hamsters, Alpha appeared to dominate in the LRT when both variants were found at similar levels in the nasal washes and URT. In contact 2 hamsters, the variant that was dominant in the URT was also dominant in the LRT (Extended Data Fig. 5c). High levels of SARS-CoV-2 replication in hamsters induced a rapid humoral immune response, as shown by serum reactivity in RBD-based ELISA in all but one of the contact hamsters (Extended Data Fig. 6a–c). We observed a twofold increase in in vitro binding affinity of recombinant trimeric S^{Alpha} to hamster ACE2 compared with S(D614G) (Extended Data Fig. 1d). These findings suggest that although S^{Alpha} has an increased binding affinity for ACE2, this factor was not predictive of the outcome of experimental infections in hamsters.

Fig. 2: Competitive replication and transmission of Alpha and wt-S^{614G} in Syrian hamsters.

 **figure 2**

Six donor hamsters were each inoculated with a TCID₅₀ of 10^{4.3}, determined by back titration and comprising a mixture of wt-S^{614G} and Alpha at a 1:1.6 ratio, determined by RT-qPCR. Donor, contact 1 and contact 2 hamsters were co-housed sequentially as shown in Extended Data Fig. 2a. Nasal washes were performed daily from 1–9 dpi and then every 2 days until 21 dpi. Pie charts show the ratio of variants detected in nasal washes at the indicated dpi. Pie chart sizes are proportional to the total number of viral genome copies per ml, as shown above or below each chart. Grey pies indicate values below the LOD. Hamster silhouettes are coloured to indicate the dominant variant (>66%) detected in the last positive sample from each hamster; a silhouette with two colours indicates that there is no dominant variant. Daggers indicate that the hamster reached the humane endpoint.

[Source data](#)

Alpha dominates wt-S^{614G} in ferrets

In a similar approach, we inoculated six donor ferrets with a mixture of wt-S^{614G} and Alpha at equivalent numbers of genome copies and monitored sequential transmission in naive contact 1 and contact 2 ferrets (Extended Data Fig. [2b](#)). Alpha rapidly became the dominant variant in nasal washes from 2 dpi with up to 10^5 gc ml⁻¹ (Fig. [3](#)). Correspondingly, the nasal concha of donor ferrets revealed high levels of replication in the nasal epithelium and up to 100-fold higher load of Alpha (up to $10^{8.2}$ gc ml⁻¹) than wt-S^{614G} (up to $10^{6.5}$ gc ml⁻¹) (Extended Data Fig. [7a](#)). Although histopathological analysis clearly indicated viral replication in the nasal epithelium of the donor ferrets (Extended Data Fig. [7b–e](#)), we did not observe severe clinical signs of infection (Extended Data Fig. [4d,e](#)). Transmission to contact 1 ferrets was detected in only two pairs of ferrets, and only one contact 1 ferret transmitted the virus to the contact 2 ferret. However, in each of these three transmission events, the Alpha variant was highly dominant and replicated to similarly high titres as in donor ferrets (Fig. [3](#)). The 3 contact ferrets with virus shedding seroconverted by 15–20 days post contact (dpc), confirming active infection (Extended Data Fig. [6d](#)).

Fig. 3: Replication and transmission of SARS-CoV-2 Alpha and wt-S^{614G} in ferrets.

 **figure 3**

Six donor ferrets were each inoculated with a TCID₅₀ of 10^{5.9}, determined by back titration and comprising a mixture of wt-S^{614G} and Alpha at a 1:1.2 ratio, determined by RT-qPCR. Donor, contact 1 and contact 2 ferrets were co-housed sequentially as shown in Extended Data Fig. 2b. Pie charts show the ratio of variants detected in nasal washes at the indicated dpi. Pie chart sizes are proportional to the total number of viral genome copies per ml, as shown above or below each chart. Grey pies indicate values below the LOD. Viral genome copies were below the LOD at 18 and 20 dpi (not

shown). Ferret silhouettes are coloured to indicate the dominant SARS-CoV-2 variant (>66%) detected in the last positive sample from each ferret.

[Source data](#)

Alpha dominates wt-S^{614G} in K18Tg mice

To assess further adaptation of Alpha to human ACE2, four hACE2-K18Tg mice, which overexpress hACE2 in respiratory epithelium¹⁸, were inoculated with a mixture of SARS-CoV-2 wt-S^{614G} and Alpha with equivalent numbers of genomic copies (Fig. 4a). Each inoculated mouse was co-housed with a contact hACE2-K18Tg mouse at 1 dpi. Alpha was dominant in the oropharyngeal samples of all four inoculated mice from 1 to 4 dpi with up to 10^6 gc ml⁻¹. The increased replicative fitness of Alpha over wt-S^{614G} was further reflected throughout the respiratory tract, with higher numbers of genome copies in nose, lungs, olfactory bulb and most brain samples at 4 dpi (Fig. 4a), and inoculated mice showed loss of body weight at 4 dpi (Extended Data Fig. 8a). A relatively high infectious dose was used to promote transmission in these experiments, and was associated with high viral load (up to 10^8 viral genome copies per sample) in the lung and brain, leading to encephalitis—as previously reported in hACE2-K18Tg mice^{19,20}. Viral loads were lower in nasal and oropharyngeal swabs from these mice, and only limited transmission was observed from these samples (two out of four contacts). None of the contact mice lost weight, but only Alpha was detectable in the lungs of contact mice at 7 dpc (Extended Data Fig. 8b).

Fig. 4: Replication of Alpha, wt-S^{Alpha}, and wt-S^{614G} in hACE2-K18Tg mice.

 **figure 4**

a, b, Two groups of four donor hACE2-K18Tg mice were inoculated with 1×10^4 PFU, determined by back titration and comprising a mixture of wt-S^{614G} (orange) and Alpha (dark blue) at a 3:1 ratio (**a**), or a mixture of wt-S^{614G} and wt-S^{Alpha} (light blue) at a 1:1 ratio (**b**). Pie charts show the ratio of variants detected in each sample at the indicated dpi. Pie chart sizes are proportional to the total number of viral genome copies per ml (swabs) or per sample (tissues), as shown below each chart. Grey pies indicate values below the LOD. Mouse silhouettes are coloured to indicate the dominant SARS-CoV-2 variant (>66%) in the last positive swab sample from the corresponding mouse; a silhouette with two colours indicates that there is no dominant variant. K18 nos. 1 to 8 denote individual hACE-K18Tg donor mice.

[Source data](#)

We performed a similar competition experiment between wt-S^{614G} and an isogenic recombinant virus expressing S^{Alpha} (wt-S^{Alpha}). We inoculated hACE2-K18Tg mice with an equal mixture of wt-S^{Alpha} and wt-S^{614G} and housed them with a contact hACE2-K18Tg mouse at 1 dpi. The replicative advantage of wt-S^{Alpha} was less clear in this experiment, and both wt-S^{Alpha} and wt-S^{614G} were present with similarly high numbers of viral genome copies in lung and brain samples (Fig. 4b). Transmission to contact mice was inefficient, and wt-S^{Alpha} was the only virus detected in lungs of contact mice at 7 dpc (Extended Data Fig. 8b). These results indicate that the S^{Alpha} spike mutations contribute to the replication advantage of Alpha over wt-S^{614G} in the URT of mice that express high levels of human ACE2.

Competition in hACE2-KI mice

To further address this question, we next used hACE2-KI homozygous mice³. In contrast to hACE2-K18Tg mice, hACE2-KI mice show physiological expression of human ACE2, with no ectopic expression of human ACE2 in the brain, and no expression of mouse ACE2, which has been shown to be permissive to the spike mutation N501Y contained in S^{Alpha}. We inoculated 4 groups of hACE2-KI mice intranasally with 10⁴ plaque-forming units (PFU) per mouse of either wt-S^{614G}, Alpha, wt-S^{Alpha} or Beta ($n = 8$ mice per group) as individual virus infections. We observed significantly higher viral genome copy numbers in mice infected with Alpha, wt-S^{Alpha} or Beta compared with wt-S^{614G} in oropharyngeal swabs at 1 dpi (Extended Data Fig. 9a). Moreover, there were significantly higher numbers of viral genome copies of Alpha and wt-S^{Alpha} in the nose at 2 dpi and in the olfactory bulb at 4 dpi compared with wt-S^{614G} and Beta (Extended Data Fig. 9b). Of note, viral titres in the nasal airways and lungs showed SARS-CoV-2 persistence at 4 dpi in 3 out of 4 mice infected with either Alpha or with wt-S^{Alpha}, but not in mice inoculated with wt-S^{614G}, whereas Beta persisted in the lungs of 2 out of 4 mice (Extended Data Fig. 9c). The apparent discrepancy between genome copy number and PFU reflects the non-homogeneous distribution of the virus in the different samples processed for each assay. We observed no difference in weight loss (Extended Data Fig. 9d) or lung histopathology score (Supplementary Table 1) between groups.

Finally, we performed competition experiments to compare the replication of the VOCs in groups of hACE2-KI mice. We observed a complete predominance of Alpha and wt-S^{Alpha} over wt-S^{614G} (Fig. 5a–c). By contrast, Beta showed reduced fitness compared with wt-S^{614G} (Fig. 5d). Together, the two mouse models support enhanced fitness of SARS-CoV-2 Alpha VOC over its progenitor wt-S^{614G} with increased replication and persistence in the URT and more systemic spread, mediated in part by changes in the Alpha S sequence.

Fig. 5: Replication of Alpha, wt-S^{Alpha}, and Beta in competition with wt-S^{614G} in hACE2-KI mice.

 figure 5

a–d, Groups of hACE2-KI male (**a, c, d**) and female (**b**) mice were inoculated with 1×10^4 PFU, determined by back titration and comprising a mixture of wt-S^{614G} and Alpha at a 3:1 ratio (**a, b**), a mixture of wt-S^{614G} and wt-S^{Alpha} at a 1:1 ratio (**c**), and a mixture of wt-S^{614G} and Beta at a 1:1.6 ratio (**d**). Pie charts show the ratio of variants

detected in each sample at the indicated dpi. Pie chart sizes are proportional to the total number of viral genome copies per ml (swabs) or per sample (tissues), as shown below each chart. Grey pies indicate values below the LOD. Mouse silhouettes are coloured to indicate the dominant SARS-CoV-2 variant (>66%) in the last positive swab sample from the corresponding mouse. KI nos. 1 to 24 denote individual hACE2-KI mice.

[Source data](#)

Discussion

Epidemiological data indicate that new SARS-CoV-2 variant lineages with specific amino acid changes have a fitness advantage over contemporary strains. VOCs such as Alpha and Beta are particularly concerning for their hypothesized ability to supersede progenitor strains and establish immune escape properties, respectively. Here we provide experimental evidence that SARS-CoV-2 Alpha has a clear replication advantage over wt-S614G in ferrets and in two humanized mouse models. Moreover, Alpha was exclusively transmitted to contact animals in competition experiments, in which ferrets and hACE2-K18Tg mice were inoculated with mixtures of Alpha and wt-S^{614G}. Because SARS-CoV-2 replicates to lower levels in ferrets and hACE2-KI mice, the inability to detect wt-S^{614G} in some samples from inoculated animals also reflects the limit of detection of the assays using PCR with reverse transcription (RT-PCR) (approximately 10³ gc ml⁻¹).

We have shown that the molecular mechanism underlying the fitness advantage of Alpha *in vivo* is largely dependent on a few changes in S, including three amino acid deletions (H69, V70 and Y144) and six substitutions (N501Y, A570D, P681H, T716I, S982A and D1118H). In hACE2-KI mice, higher genome copies and/or titres of Alpha and wt-S^{Alpha} compared with wt-S^{614G} were found in the URT (oropharynx and nose) and olfactory bulb. Increased replication and transmission of wt-S^{Alpha} over wt-S^{614G} were also evident in hACE2-K18Tg mice. Transmission events are rare in mice; however, we observed transmission of Alpha and wt-S^{Alpha} in 50% of the contact hACE2-K18Tg mice and no detection of wt-S^{614G} in any contact mouse. *In vitro*, Alpha S mutations increased its affinity for hamster and human ACE2 by twofold and fourfold, respectively, indicating an overall improvement in binding abilities rather than a specialization towards human ACE2.

Beta showed a higher binding affinity for human ACE2 than its progenitor wt-S^{614G} and an equal level of replication to Alpha and wt-S^{614G} in single infections of AEC cultures and in hACE2-KI mice. However, Beta replication was outcompeted in direct competition with wt-S^{614G} *in vitro* and in hACE2-KI mice. In hamsters, wt-S^{614G} and Alpha also outcompeted Beta in relation to replication and transmission to contact

animals, in which Beta was outnumbered by one or two orders of magnitude. This reduced fitness was also evident in previous experiments in K18-hACE2 mice²¹. The relative reduced intrinsic fitness of Beta in immunologically naive hosts supports the hypothesis that the epidemiological advantage of Beta may be principally owing to immune escape, as indicated by reduced efficiency in serum neutralization tests¹⁶. In convalescent or vaccinated populations, the immune escape advantage of Beta may prove to be sufficient to compensate its reduced intrinsic fitness and explains, for example, the low prevalence of this variant in regions with a mainly naive population.

Alpha and wt-S^{614G} exhibited similar replication and transmission in hamsters, a model with very high susceptibility and replication efficacy, in which the impact of a marginally fitter SARS-CoV-2 variant may not become apparent. Indeed, efficient simultaneous transmission of both variants to contact hamsters was observed in association with high viral loads in infected animals. In models supporting high replication, such as human AEC cultures and hamsters, only major improvements in replication and transmission can be detected when the variants compared already have a high fitness. By contrast, in ferrets and mouse models—in which SARS-CoV-2 replication is overall less efficient—VOCs with modestly enhanced replication and transmission can be identified. The similar replication and transmission efficacies in hamsters are in line with recent publications using VOCs in the hamster model²².

The basal rate of replication is an important factor in the assertion of a variant over a contemporary variant in a naive population. Some individuals with higher bioaerosol exhalation levels can initiate disproportionate numbers of transmission events, possibly because of higher viral load in the URT, and are therefore called ‘superspreaders’²³. The hamster model might thus resemble the human superspread scenario, since there is no clear indication of a specific predominance in transmission between two SARS-CoV-2 variants with high fitness levels, such as wt-S^{614G} and Alpha. However, we did not perform strict aerosol transmission studies, so this remains only a proposition. The ferret and hACE2-KI models are more restricted in that infection is predominantly in the URT. Therefore, these models more closely mimic the situation in humans, in which infections are predominantly mild. Although the rate of transmission was not high overall (3 out of 8 pairs in ferrets, and 4 out of 8 pairs in hACE2-K18Tg mice), the almost exclusive transmission of Alpha relative to wt-S^{614G} mirrored increased transmission of Alpha in the human population; Alpha has been responsible for more than 90% of infections in most countries in Europe²⁴.

Overall, our study demonstrates that multiple complementary models are necessary to comprehensively evaluate different aspects of human SARS-CoV-2 infection and the impact of emerging VOCs on the course of the ongoing pandemic. The hamster and ferret provide complementary models for transmission efficiency. The mouse models used here may become critical for VOCs demonstrating higher specificity for binding

to human ACE2 relative to those from other species. Together, our results show the clear fitness advantage of Alpha and a concomitant disadvantage of Beta, in line with the observed epidemiological predominance of Alpha in the context of a relatively naive population. Notably and reassuringly, despite the apparent fitness differences of these VOCs, there is no indication of different pathologies.

Methods

Cell lines

Vero E6 cells (ATCC CRL-1586) (provided by D. Muth, M. Müller and C. Drosten) or Vero-TMPRSS2²⁵ (provided by S. Pöhlmann) were propagated in Dulbecco's Modified Eagle Medium-GlutaMAX supplemented with 1 mM sodium pyruvate, 10% (v/v) heat-inactivated fetal bovine serum (FBS), 100 µg ml⁻¹ streptomycin, 100 IU ml⁻¹ penicillin, 1% (w/v) nonessential amino acids and 15 mM HEPES (Gibco). Cells were maintained at 37 °C in a humidified incubator with 5% CO₂.

Viruses

Viruses are listed in Extended Data Table 1 together with the corresponding in vitro and in vivo experiments in which they were used. Specific amino acid changes are shown schematically in Extended Data Fig. 10. Contemporary clinical isolates from the B.1.160 (S^{D614G}) (EPI_ISL_414019), Alpha (EPI_ISL_2131446, EPI_ISL_751799 (L4549)) and Beta (EPI_ISL_803957 (L4550)) were isolated and minimally passaged on Vero E6 cells. Beta (EPI_ISL_981782) was initially isolated on A549 cells expressing human ACE2 before passaging on Vero E6 cells. SARS-CoV-2 Alpha (L4549) and Beta (L4550)²¹ were received from the Robert-Koch-Institut Berlin, Germany. Isogenic variants with the Alpha spike (wt-S^{Alpha}) or individual Alpha spike mutations were introduced into a wild-type SARS-CoV-2 ‘Wuhan’ backbone strain comprising the D614G amino acid change (wt-S^{D614G}), as described^{3,26}. Isogenic viruses were grown on Vero-TMPRSS2 cells after one passage on human bronchial airway epithelial cells. All viruses were verified by performing whole-genome next generation sequencing (NGS). For SARS-CoV-2 Alpha (L4549, SARS-CoV-2 B.1.1.7 NW-RKI-I-0026/2020 passage 3), one silent mutation in the ORF1a (sequence position 11741) was determined (C to T with 27% T, 57% strand bias). For SARS-CoV-2 Beta (L4550, available under ENA study accession number MZ433432), one nucleotide exchange was detected (A12022C) resulting in the amino acid exchange D3923A in ORF1a and one SNP at sequence position 11730 (C to T with 41%, stand bias 52%).

For all in vivo virus competition experiments, we generated inoculum mixtures aiming for a 1:1 ratio of each variant based on virus stock titres. The reported mixture inoculum titres are based on back-titration of the inoculum mixtures and the indicated ratio of each variant was determined by standard RT-qPCR. SARS-CoV-2 wt-S^{614G} (PRJEB45736; wt-S614G ID#49 vial 2) and SARS-CoV-2 Beta (L4550) were used to inoculate hamsters in the wt-S^{614G} versus Beta study; SARS-CoV-2 Alpha (L4549), and SARS-CoV-2 Beta (L4550) were used for inoculation in the Alpha versus Beta hamster study. SARS-CoV-2 wt-S^{614G}, wt-S^{Alpha}, Alpha (L4549) and Beta (L4550) were used to inoculate hACE2 humanized mice in all single virus or mixed virus competition experiments.

Next-generation sequencing

NGS was used to verify the sequence of isolates and isogenic clones prior to experimentation. RNA was extracted using the RNAdvance Tissue kit (Beckman Coulter) and the KingFisher Flex System (Thermo Fisher Scientific). Subsequently, RNA was transcribed into cDNA and sequencing libraries were generated as described²⁷ and were sequenced using the Ion Torrent S5XL Instrument (ThermoFisher). Samples with C_t values >20 for SARS-CoV-2 were additionally treated with RNA baits (myBaits, Arbor Biosciences) for SARS-CoV-2 enrichment before sequencing²⁸. Sequence datasets were analysed by reference mapping with the Genome Sequencer Software Suite (version 2.6, Roche), default software settings for quality filtering and mapping using EPI_ISL_414019 (Alpha), EPI_ISL_2131446 (Alpha) and EPI_ISL_981782 (Beta) as references. To identify potential single nucleotide polymorphisms in the read data, the variant analysis tool integrated in Geneious Prime (2019.2.3) was applied (default settings).

AEC cultures

Human nasal AEC cultures were purchased from Epithelix (EP02MP Nasal MucilAir, pool of 14 donors). Maintenance of primary nasal AEC cultures were performed according to manufacturer's guidelines. Individual SARS-CoV-2 infections with contemporary virus isolates were conducted at either 33 °C or 37 °C as described elsewhere²⁹ using a multiplicity of infection (MOI) of 0.02, whereas all competition experiments and replication kinetics were performed with an MOI of 0.005 as described³⁰. Quantification of viral load of individual SARS-CoV-2 infections with contemporary virus isolates was performed using the NucliSens easyMAG (BioMérieux) and RT-qPCR targeting the *E* gene of SARS-CoV-2 as described^{31,32}. In competition experiments, nucleic acids were extracted using the Quick-RNA Viral 96 kit (Zymo research) and RT-qPCR primers and probe sequences are shown in

Extended Data Table 2. The viral replication of individual isogenic variants was monitored by plaque assay.

Plaque titration assay

Viruses released into the apical compartments were titrated by plaque assay on Vero E6 cells as described^{30,33}. In brief, 2×10^5 cells per ml were seeded in 24-well plates 1 day prior to titration and inoculated with tenfold serial dilutions of virus solutions. Inocula were removed 1 h post-infection and replaced with overlay medium consisting of DMEM supplemented with 1.2% Avicel (RC-581, FMC biopolymer), 15 mM HEPES, 5 or 10% heat-inactivated FBS, 100 $\mu\text{g ml}^{-1}$ streptomycin and 100 IU ml^{-1} penicillin. Cells were incubated at 37 °C, 5% CO₂ for 48 h, fixed with 4% (v/v) neutral buffered formalin, and stained with crystal violet.

Protein expression, purification and bio-layer interferometry assay

SARS-CoV-2 S protein expression plasmids were constructed to encode the ectodomain of S protein S(D614G) or S^{Alpha} (residues 1–1208, with a mutated furin cleavage site and K986P/V987P substitutions) followed by a T4 fold on the trimerization domain and a polyhistidine purification tag. ACE2 protein (human, hamster or ferret) expression plasmids were constructed to encode the ectodomain of ACE2 followed by a human IgG1 Fc purification tag. The recombinant proteins were expressed using the Expi293 Expression system (ThermoFisher Scientific) and purified with HisTrap FF columns (for polyhistidine-tagged spike proteins) or with HiTrap Protein A column (for Fc-tagged ACE2 proteins) in FPLC (Cytiva) system. Recombinant proteins were further purified with Superose 6 Increase 10/300 GL column (Cytiva) as needed.

Binding affinity between the trimeric spike and dimeric ACE2 was evaluated using an Octet RED96e instrument at 30 °C with a shaking speed of 1,000 rpm (ForteBio). Anti-human IgG Fc biosensors (ForteBio) were used. Following 20 min of pre-hydration of anti-human IgGFc biosensors and 1 min of sensor check, 7.5 nM of human ACE2–Fc, 7.5 nM of hamster ACE2–Fc in 10× kinetic buffer (ForteBio) were loaded onto the surface of anti-human IgG Fc biosensors for 5 min. After 1.5 min of baseline equilibration, 5 min of association was conducted at 10–100 nM S(D614G), S^{Alpha} or S^{Beta}, followed by 5 min of dissociation in the same buffer, which was used for baseline equilibration. The data were collected using ForteBio Data Acquisition Software 12.0.1 and corrected by subtracting signal from the reference sample and a 1:1 binding model with global fit was used for determination of affinity constants.

Animal experiment ethics declarations

All ferret and hamster experiments were evaluated by the responsible ethics committee of the State Office of Agriculture, Food Safety, and Fishery in Mecklenburg–Western Pomerania (LALLF M-V) and gained governmental approval under registration number LVL MV TSD/7221.3-1-004/21. Mouse studies were approved by the Commission for Animal Experimentation of the Cantonal Veterinary Office of Bern and conducted in compliance with the Swiss Animal Welfare legislation and under license BE-43/20.

Hamster studies

Six Syrian hamsters (*Mesocricetus auratus*) (Janvier Labs) were inoculated intranasally under a brief inhalation anaesthesia with a 70 µl mixture of two SARS-CoV-2 VOCs (wt-S^{614G} and Alpha mixture, wt-S^{614G} and Beta mixture, or Alpha and Beta mixture). Each inoculum was back-titrated and ratios of each variant were determined by RT-qPCR. The wt-S^{614G} and Alpha mixture held a 1:1.6 ratio with TCID₅₀ of 10^{4.3} per hamster, the wt-S^{614G} versus Beta mixture held a 1:3.8 ratio with TCID₅₀ of 10^{4.25} per hamster, and the Alpha versus Beta mixture held a 1.8:1 ratio with TCID₅₀ of 10^{5.06} per hamster.

Inoculated donor hamsters were isolated in individually ventilated cages for 24 h. Thereafter, contact hamster 1 was co-housed with each donor, creating six donor–contact 1 pairs (Extended Data Fig. 2a). The housing of each hamster pair was strictly separated in individual cage systems to prevent spillover between different pairs. At 4 dpi, the individual donor hamsters (inoculated animal) were euthanized. To simulate a second transmission cycle, the original contact hamsters (referred to as contact 1) were commingled with a further six naive hamsters (referred to as contact 2), which equates to another six contact 1 and contact 2 pairs (Extended Data Fig. 2a). These pairs were co-housed until the end of the study at 21 dpi. Because the first contact hamster (cage 6) in the competition trial wt-S^{614G} versus Alpha, died at 2 dpc, the second contact hamster for this cage was also co-housed with the donor hamster; thus the first and second contact hamsters in this cage were labelled contact 1a and contact 1b, respectively. To enable sufficient contact between the donor hamster and contact 1b hamster, which was commingled routinely on 4 dpi, the donor hamster was euthanized at 7 dpi (instead of at 4 dpi), when it reached the humane end-point criterion for bodyweight (below 80% of 0 dpi body weight).

Viral shedding was monitored by nasal washes in addition to a daily physical examination and body weighing routine. Nasal wash samples were obtained under a short-term isoflurane anaesthesia from individual hamsters by administering 200 µl PBS to each nostril and collecting the reflux. Animals were sampled daily from 1 dpi to 9 dpi, and then every other day until 21 dpi. Under euthanasia, serum samples and

an organ panel comprising representative URT and LRT tissues were collected from each hamster. All animals were observed daily for signs of clinical disease and weight loss. Hamsters reaching the humane endpoint, that is, falling below 80% of the initial body weight relative to 0 dpi, were humanely euthanized.

Ferret studies

Similar to the hamster study, 12 ferrets (six donor ferrets and six transmission 1 ferrets) from the FLI in-house breeding were housed pairwise in strictly separated cages to prevent spillover contamination. Of these, six ferrets were inoculated with an equal 250 µl mixture of SARS-CoV-2 wt-S^{614G} and Alpha. The inoculum was back-titrated and the ratio of each variant was determined by RT-qPCR. The wt-S^{614G} versus Alpha mixture held a 1:1.2 ratio with 10^{5.875} TCID₅₀ distributed equally into each nostril of donor ferrets. Ferrets were separated for the first 24 h following inoculation. Subsequently, the ferret pairs were co-housed again, allowing direct contact of donor to contact 1 ferrets. All ferrets were sampled via nasal washes with 750 µl PBS per nostril under a short-term inhalation anaesthesia. Donor ferrets were sampled until euthanasia at 6 dpi, which was followed by the introduction of one additional naive contact 2 ferret per cage ($n = 6$), resulting in a 1:1 pairwise setup with contact 1 and contact 2 ferrets (Extended Data Fig. [2b](#)). All ferrets, which were in the study group on the respective days, were sampled on the indicated days. Bodyweight, temperature and physical condition of all ferrets were monitored daily throughout the experiment. URT and LRT organ samples, as well as blood samples of all ferrets were taken at respective euthanasia time points.

Full autopsy was performed on all animals under BSL3 conditions. The lung, trachea and nasal conchae were collected and fixed in 10% neutral-buffered formalin for 21 days. The nasal atrium, decalcified nasal turbinates (cross-sections every 3–5 mm), trachea and all lung lobes were trimmed for paraffin embedding. Based on PCR results, tissue sections (3 µm) of all donors (day 6) and one recipient (no. 8, day 20) were cut and stained with haematoxylin and eosin for light microscopical examination. Immunohistochemistry was performed using an anti-SARS nucleocapsid antibody (Novus Biologicals NB100-56576, dilution 1:200) according to standardized avidin–biotin–peroxidase complex-method producing a red labelling and haematoxylin counterstain. For each immunohistochemistry staining, positive control slides and a negative control for the primary antibodies were included. Histopathology was performed on at least five consecutive tissue samples per animal, yielding comparable results in all cases. Lung tissue pathology was evaluated according to a detailed score sheet developed by Angele Breithaupt (DipECVP) (Supplementary Table [2](#)). Evaluation and interpretation was performed by board-certified veterinary pathologists (DiplECVP) (AB, IBV).

Mouse studies

hACE2-KI mice (*B6.Cg-Ace2^{tm1(ACE2)Dwnt}*) and hACE2-K18Tg mice (*Tg(K18-hACE2)2Prlmn*) were described previously^{3,18}. All mice were produced at the specific-pathogen-free facility of the Institute of Virology and Immunology (Mittelhäusern), where they were maintained in individually ventilated cages (blue line, Tecniplast), with 12-h:12-h light:dark cycle, 22 ± 1 °C ambient temperature and $50 \pm 5\%$ humidity, autoclaved food and acidified water. At least 7 days before infection, mice were placed in individually HEPA-filtered cages (IsoCage N, Tecniplast). Mice (10 to 12 weeks old) were anaesthetized with isoflurane and infected intranasally with 20 µl per nostril with the virus inoculum described in the results section. One day after inoculation, infected hACE2-K18Tg mice were placed in the cage of another hACE2-K18Tg contact mouse. Mice were monitored daily for bodyweight loss and clinical signs. Oropharyngeal swabs were collected under brief isoflurane anaesthesia using ultrafine sterile flock swabs (Hydraflock, Puritan, 25-3318-H). The tips of the swabs were placed in 0.5 ml of RA1 lysis buffer (Macherey-Nagel, 740961) supplemented with 1% β-mercaptoethanol and vortexed. At 2 or 4 dpi, mice were euthanized, and organs were aseptically dissected. Systematic tissue sampling was performed as detailed previously³.

Animal specimens work up, viral RNA detection and quantification

Organ samples from ferrets and hamsters were homogenized in a 1 ml mixture composed of equal volumes of Hank's balanced salts MEM and Earle's balanced salts MEM containing 2 mM l-glutamine, 850 mg l^{-1} NaHCO₃, 120 mg l^{-1} sodium pyruvate and 1% penicillin–streptomycin) at 300 Hz for 2 min using a Tissuelyser II (Qiagen) and centrifuged to clarify the supernatant. Organ samples from mice were either homogenized in 0.5 ml of RA1 lysis buffer supplemented with 1% β-mercaptoethanol using a Bullet Blender Tissue Homogenizer (Next-Advance) or in Tube M (Miltenyi Biotech, 130-096-335) containing 1 ml of DMEM using a gentleMACS Tissue Dissociator (Miltenyi Biotech). Nucleic acid was extracted from 100 µl of the nasal washes or 200 µl mouse oropharyngeal swabs after a short centrifugation step or 100 µl of organ sample supernatant using the NucleoMag Vet kit (Macherey Nagel). Nasal washes, oropharyngeal swabs, and organ samples were tested by RT–qPCR analysis for the ratio of the two different viruses used for inoculation, by applying two different assays, each of them specific for one variant: either the wt-S^{614G}, Alpha or Beta variant (Extended Data Tables 2, 3). Viral RNA copies in swabs and organs in studies using a single variant inoculum in mice were determined using the E protein RT–qPCR exactly as described³.

Four specific RT–qPCR assays for SARS-CoV-2 wt-S^{614G}, Alpha and Beta were designed based on the specific genome deletions within ORF1 and the *S* gene (Extended Data Table 2). Here, virus-specific primers were used to achieve a high analytical sensitivity (less than 10 genome copies per μ l template) of the PCR assays, and in samples with a high genome load of the non-matching virus.

The RT–qPCR reaction was prepared using the qScript XLT One-Step RT–qPCR ToughMix (QuantaBio) (hamsters and ferrets) or the AgPath-ID One-Step RT–PCR (ThermoFisher Scientific) (hACE2-K18Tg and hACE2-KI mice) in a volume of 12.5 μ l including 1 μ l of the respective FAM mix and 2.5 μ l of extracted RNA. The reaction was performed for 10 min at 50 °C for reverse transcription, 1 min at 95 °C for activation, and 42 cycles of 10 s at 95 °C for denaturation, 10 s at 60 °C for annealing and 20 s at 68 °C for elongation. Fluorescence was measured during the annealing phase. RT–qPCRs were performed on a BioRad real-time CFX96 detection system (Bio-Rad) (hamsters and ferrets) or an Applied Biosystems 7500 Real-Time PCR System (ThermoFisher Scientific) (mice). Validation work was performed by comparison with established protocols (https://www.who.int/docs/default-source/coronavirus/eal-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2 and ref. 31).

Serological tests of hamsters and ferrets

Serum samples from the wt-S^{614G} versus Alpha, wt-S^{614G} versus Beta, and Alpha versus Beta co-inoculated hamsters and ferrets were tested by ELISA for sero-reactivity against the RBD domain³⁴ using a Tecan i-control 2014 1.11 plate reader and data was analysed using Microsoft Excel 16.0. All samples were generated at the time point of euthanasia of the individual animal.

Statistical analysis

Statistical analysis was performed using GraphPad Prism 8 or R³⁵ (version 4.1), using the packages tidyverse³⁶ (v1.3.1), ggpubr (v0.4.0) and rstatix (v.0.7.0). Unless noted otherwise, the results are expressed as mean \pm s.d. Two-way analysis of variance (ANOVA) with Tukey honest significance differences post hoc test was used to compare competition results at different time points after infection in vitro. One-way ANOVA with Tukey's multiple comparisons test was used to compare viral genome copies or titres at different time points post infection in individual virus mouse infection studies. Significance was defined as $P < 0.05$.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Sequence data are available on the NCBI Sequence Read Archive (SRA) under the accession numbers [PRJEB45736](#) and [PRJNA784099](#), or in GenBank under the accession numbers [MT108784](#), [MZ433432](#), [OL675863](#), [OL689430](#) and [OL689583](#) as shown in Extended Data Table 1. [Source data](#) are provided with this paper.

References

1. Harvey, W. T. et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
2. Tao, K. et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* **22**, 757–773 (2021).
3. Zhou, B. et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **592**, 122–127 (2021).
4. 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.e819 (2020).
5. Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
6. Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
7. Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
8. Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).
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. 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.e2587 (2021).

11. Weber, S., Ramirez, C. M., Weiser, B., Burger, H. & Doerfler, W. SARS-CoV-2 worldwide replication drives rapid rise and selection of mutations across the viral genome: a time-course study—potential challenge for vaccines and therapies. *EMBO Mol. Med.* **13**, e14062 (2021).
12. Galloway, S. E. et al. Emergence of SARS-CoV-2 B.1.1.7 lineage—United States, December 29, 2020–January 12, 2021. *Morb. Mortal. Wkly Rep.* **70**, 95–99 (2021).
13. Liu, Y. et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* <https://doi.org/10.1038/s41586-021-04245-0> (2021).
14. Kemp, S. et al. Recurrent emergence and transmission of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
15. Cele, S. et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **593**, 142–146 (2021).
16. Hoffmann, M. et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* **184**, 2384–2393.e2312 (2021).
17. 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).
18. McCray, P. B., Jr et al. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 813–821 (2007).
19. Yinda, C. K. et al. K18-hACE2 mice develop respiratory disease resembling severe COVID-19. *PLoS Pathog.* **17**, e1009195 (2021).
20. Zheng, J. et al. COVID-19 treatments and pathogenesis including anosmia in K18-hACE2 mice. *Nature* **589**, 603–607 (2021).
21. Hoffmann, D. et al. CVnCoV and CV2CoV protect human ACE2 transgenic mice from ancestral B BavPat1 and emerging B.1.351 SARS-CoV-2. *Nat. Commun.* **12**, 4048 (2021).
22. Abdelnabi, R. et al. Comparing infectivity and virulence of emerging SARS-CoV-2 variants in Syrian hamsters. *EBioMedicine* **68**, 103403 (2021).
23. Edwards, D. A. et al. Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proc. Natl Acad. Sci. USA* **118**, e2021830118 (2021).

24. Variants of concern. *European Centre for Disease Prevention and Control* <https://www.ecdc.europa.eu/en/covid-19/variants-concern> (accessed 3 June 2021).
25. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e278 (2020).
26. Thi Nhu Thao, T. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).
27. Wylezich, C., Papa, A., Beer, M. & Hoper, D. A versatile sample processing workflow for metagenomic pathogen detection. *Sci. Rep.* **8**, 13108 (2018).
28. Wylezich, C. et al. Next-generation diagnostics: virus capture facilitates a sensitive viral diagnosis for epizootic and zoonotic pathogens including SARS-CoV-2. *Microbiome* **9**, 51 (2021).
29. Essaidi-Laziosi, M. et al. Propagation of respiratory viruses in human airway epithelia reveals persistent virus-specific signatures. *J. Allergy Clin. Immunol.* **141**, 2074–2084 (2018).
30. V'Kovski, P. et al. Disparate temperature-dependent virus-host dynamics for SARS-CoV-2 and SARS-CoV in the human respiratory epithelium. *PLoS Biol.* **19**, e3001158 (2021).
31. Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, 2000045 (2020).
32. Baggio, S. et al. SARS-CoV-2 viral load in the upper respiratory tract of children and adults with early acute COVID-19. *Clin. Infect. Dis.* **73**, 148–150 (2020).
33. Jonsdottir, H. R. & Dijkman, R. Characterization of human coronaviruses on well-differentiated human airway epithelial cell cultures. *Methods Mol. Biol.* **1282**, 73–87 (2015).
34. Wernike, K. et al. Multi-species ELISA for the detection of antibodies against SARS-CoV-2 in animals. *Transbound. Emerg. Dis.* **68**, 1779–1785 (2020).
35. R: A language and environment for statistical computing (R Core Team, 2021).
36. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

Acknowledgements

We thank F. Klipp, D. Fiedler, C. Lipinski, S. Kiepert, K. Sliz and Daniel Brechbühl for animal care; M. Lange, C. Korthase, P. Zitzow, S. Schuparis, G. Cadau, P. Valenti, B. Lemaitre, C. Tougne, P. Fontannaz, P. Sattonnet and C. Alvarez for technical assistance; and M. Schmolke, B. Mazel-Sanchez and F. Abdul for providing A549-hACE2 cells. This work was supported by grants from the Swiss National Science Foundation (SNSF), grants no. 31CA30_196062 (C.B. and R.D.), 31CA30_196644 (V.T., I.A.E. and R.D.), 310030_173085 (V.T.), 310030_179260 (R.D.), 196383 (I.A.E.); the European Commission, Marie Skłodowska-Curie Innovative Training Network ‘HONOURS’, grant agreement no. 721367 (V.T. and R.D.); the European Union Project ReCoVer, grant no. GA101003589 (C. Drosten); Core funds of the University of Bern (V.T. and R.D.); Core funds of the German Federal Ministry of Food and Agriculture (M. Beer); the Deutsche Forschungsgemeinschaft (DFG), project no. 453012513 (M. Beer); the Horizon 2020 project ‘VEO’, grant agreement no. 874735 (M. Beer); COVID-19 special funds from the Swiss Federal Office of Public Health and the Swiss Federal Office of Food Safety and Veterinary Affairs (A.S. and V.T.); the Fondation Ancrage Bienfaisance du Groupe Pictet (I.A.E.); the Fondation Privée des Hôpitaux Universitaires de Genève (I.A.E.); and the German Ministry of Research, VARIPath, grant no. 01KI2021 (V.M.C.).

Author information

Author notes

1. These authors contributed equally: Lorenz Ulrich, Nico Joel Halwe, Adriano Taddeo, Nadine Ebert
2. These authors jointly supervised this work: Ronald Dijkman, Donata Hoffmann, Volker Thiel, Martin Beer, Charaf Benarafa

Affiliations

1. Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany

Lorenz Ulrich, Nico Joel Halwe, Jacob Schön, Bernd Hoffmann, Anna Kraft, Claudia Wylezich, Donata Hoffmann & Martin Beer

2. Institute of Virology and Immunology, Mittelhäusern, Switzerland

Adriano Taddeo, Nadine Ebert, Christelle Devisme, Bettina Salome Trüeb, Aurélie Godel, Jenna N. Kelly, Inês Berenguer Veiga, Lisa Thomann, Artur Summerfield, Volker Thiel & Charaf Benarafa

3. Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Adriano Taddeo, Nadine Ebert, Christelle Devisme, Bettina Salome Trüeb, Aurélie Godel, Jenna N. Kelly, Inês Berenguer Veiga, Lisa Thomann, Artur Summerfield, Volker Thiel & Charaf Benarafa

4. Institute of Veterinary Bacteriology, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Bettina Salome Trüeb, Fabien Labroussaa & Jörg Jores

5. Institute for Infectious Diseases, University of Bern, Bern, Switzerland

Manon Wider, Laura Laloli, Mitra Gultom, Monika Gsell & Ronald Dijkman

6. CDC COVID-19 Emergency Response, Centers for Disease Control and Prevention, Atlanta, GA, USA

Xiaoyu Fan, Brenda M. Calderon, Sarah Osman, Bin Zhou & David E. Wentworth

7. Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Meriem Bekliz, Manel Essaidi-Laziosi, Kenneth Adea & Isabella Anne Eckerle

8. Charité–Universitätsmedizin Berlin, Institute of Virology, Berlin, Germany

Marie Luisa Schmidt, Daniela Niemeyer, Victor Max Corman & Christian Drosten

9. German Centre for Infection Research (DZIF), Berlin, Germany

Daniela Niemeyer, Victor Max Corman & Christian Drosten

10. Graduate School for Biomedical Science, University of Bern, Bern, Switzerland

Laura Laloli & Mitra Gultom

11. Department of Experimental Animal Facilities and Biorisk Management,
Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany
- Angele Breithaupt
12. Centre for Vaccinology, Department of Pathology and Immunology, University of
Geneva, Geneva, Switzerland
- Benjamin Meyer & Christiane S. Eberhardt
13. Division of General Paediatrics, Department of Woman, Child and Adolescent
Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland
- Christiane S. Eberhardt
14. Center for Vaccinology, Geneva University Hospitals, Geneva, Switzerland
- Christiane S. Eberhardt
15. Division of Infectious Disease, Geneva University Hospitals, Geneva,
Switzerland
- Isabella Anne Eckerle
16. Division of Laboratory Medicine, Laboratory of Virology, Geneva University
Hospitals, Geneva, Switzerland
- Isabella Anne Eckerle

Contributions

Conceptualization: D.H., M. Beer, V.T. and C.B. Data curation: L.U., N.J.H., A.T., N.E., J.S., C. Devisme, B.Z. and R.D. Funding acquisition: A.S., I.A.E., D.E.W., R.D., V.T., M. Beer and C.B. Investigation: L.U., N.J.H., A.T., N.E., J.S., C. Devisme, B.S.T., B.H., M.W., X.F., M. Bekliz, M.E.-L., M.L.S., D.N., V.M.C., A.K., A.G., L.L., J.N.K., B.M.C., A.B., C.W., I.B.V., M. Gultom, S.O., B.Z., K.A., B.M., C.S.E., L.T., M. Gsell, R.D. and D.H. Methodology: B.H., A.B., F.L. and J.J. Supervision: I.A.E., C. Drosten, R.D., D.H., V.T., M. Beer and C.B. Visualization: L.U., N.J.H., A.T., J.S., C. Devisme and R.D. Writing, original draft: L.U., N.J.H., A.T., R.D., D.H., M. Beer and C.B. Writing, review and editing: L.U., N.J.H., A.B., L.T., R.D., D.H., V.T., M. Beer and C.B.

Corresponding authors

Correspondence to [Volker Thiel](#), [Martin Beer](#) or [Charaf Benarafa](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Stanley Perlman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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 ACE2 receptor binding and replication kinetics of SARS-CoV-2 VOC in vitro.](#)

(a) Affinity between spike (S^{614G} , S^{Alpha} , and S^{Beta}) protein trimers and hACE2 dimers determined by Bio-layer interferometry. **(b)** Viral replication kinetics of SARS-CoV-2 Alpha, Beta, and wt- S^{614G} (MOI 0.02) at 33 °C and 37 °C in primary human nasal airway epithelial cell (AEC) cultures. **(c)** Viral replication kinetics of pairwise competition assays in primary nasal AEC cultures at 33 °C (MOI 0.005). **(b, c)** Data are presented as individual points with mean (line) and standard deviation; **(b)** $n = 2$ (Alpha and Beta), $n = 4$ (wt- S^{614G}), **(c)** $n = 3$ independent biological replicates. **(c)** P-values were determined by two-way ANOVA and Tukey Honest Significant Differences (HSD) post-hoc test. **(d)** Affinity between spike (S^{614G} , S^{Alpha}) protein trimers with hamster ACE2 determined by Bio-layer interferometry. **(a, d)** ACE2 with IgG1 Fc tag were loaded on anti-human IgG Fc biosensors and binding kinetics were conducted using indicated concentrations of spike trimers. Data is representative of 3 independent experiments.

[Source data](#)

Extended Data Fig. 2 Experimental workflow of competitive transmission experiments in Syrian hamsters and ferrets.

(a) Timeline of the hamster experiments. Six intranasally inoculated donor hamsters each were co-housed with one naïve contact hamster (1 dpi), building six respective donor-contact I pairs. At 4 dpi, the donor hamsters were euthanized and the initial contact hamsters I were co-housed with one additional hamster (Contact II). (b) Timeline of the ferret experiment. The scheme was generated with BioRender (<https://biorender.com/>).

Extended Data Fig. 3 Competitive transmission between Alpha and Beta in Syrian hamsters.

Six donor hamsters were each inoculated with $10^{5.06}$ TCID₅₀ determined by back titration and composed of a mixture of Alpha (dark blue) and Beta (green) at 1.8:1 ratio determined by RT-qPCR. Donor hamsters, contact I and II hamsters were co-housed sequentially as shown in Extended Data Fig. 2a. Nasal washings were performed daily from 1–9 dpi and afterwards every two days until 21 dpi. Pie chart colors illustrate the ratio of variants detected in nasal washings at the indicated dpi. Pie chart sizes are proportional to the total viral genome copies reported above or below respective pies. Grey pies indicate values below the LOD (<103 viral genome copies per sample). Hamster silhouettes are colored according to the dominant variant (>66%) detected in the latest sample of each animal. † indicate that the corresponding animal reached the humane endpoint.

[Source data](#)

Extended Data Fig. 4 Clinical features of hamsters and ferrets.

(a–c) Syrian hamsters were inoculated with comparable genome equivalent mixture of either wt-S^{614G} and Beta (a), Alpha and Beta (b), or wt-S^{614G} and Alpha (c). In hamsters, body weight was monitored daily until 13 dpi, afterwards every two days until 21 dpi and plotted relative to bodyweight of day 0. The dotted line indicates the humane endpoint criterion of 20% body weight loss from initial bodyweight at which hamsters were promptly euthanized for animal welfare reasons. (d, e) Ferrets were inoculated intranasally with an equal mixture of wt-S^{614G} and Alpha. Body weight (d) and temperature (e) were monitored daily in ferrets until 12 dpi, and afterwards every 2 days. Grey dotted lines in e indicate the physiologic range for body temperature in ferrets.

[Source data](#)

Extended Data Fig. 5 Viral genome load in upper (URT) and lower (LRT) respiratory tract tissues of Syrian hamsters in the competitive transmission experiment between SARS-CoV-2 VOCs.

(a–c) Syrian hamsters were inoculated with comparable genome equivalent mixture of either wt-S^{614G} and Beta **(a)**, Alpha and Beta **(b)**, or wt-S^{614G} and Alpha **(c)**. Absolute quantification was performed by RT–qPCR analysis of tissue homogenates of donor, contact I and contact II hamsters in relation to a set of defined standards. Tissue samples were collected at euthanasia (Euth.). Pie chart colors illustrate the ratio of variants detected in each sample at the indicated dpi or days post contact (dpc). Pie chart sizes are proportional to the total viral genome copies reported below. Grey pies indicate values below the LOD (<10³ viral genome copies per sample).

[Source data](#)

Extended Data Fig. 6 Indirect ELISA against the RBD of SARS-CoV-2.

Sera of donor hamsters **(a, b, c)** and ferrets **(d)** inoculated with the indicated SARS-CoV-2 VOC mixtures and sera of contact I and II animals were collected at their respective experimental endpoints . All sera were tested for specific reactivity against the SARS-CoV-2 RBD-SD1 domain (wt-S amino acids 319-519).

[Source data](#)

Extended Data Fig. 7 Viral genome load in upper (URT) and lower (LRT) respiratory tract tissue of ferrets in the competitive transmission experiment between SARS-CoV-2 Alpha and wt-S^{614G}.

(a) Absolute quantification was performed by RT–qPCR analysis of tissue homogenates of donor, contact I and contact II ferrets in relation to a set of defined standards. Tissue samples were collected at euthanasia (Euth.). Pie chart colors illustrate the ratio of variants detected in each sample at the indicated dpi or dpc. Pie chart sizes are proportional to the total viral genome copies reported below. Grey pies indicate values below the LOD (<10³ viral genome copies per sample). **(b–e)** Representative micrographs of hematoxylin and eosin staining of 3 µm sections of nasal conchae of donor ferrets (n = 6) 6 dpi with wt-S^{614G} and Alpha. Micrographs are representative of 5 consecutive tissue samples of each animal. Insets show immunohistochemistry staining of SARS-CoV-2 with anti-SARS nucleocapsid antibody with hematoxylin counterstain. The respiratory **(b, c)** and olfactory **(d, e)** nasal mucosa exhibited rhinitis with varying severity. Lesion-associated antigen was found in ciliated cells of the respiratory epithelium **(b, c)** and in sustentacular cells of

the olfactory epithelium (**d**, **e**) in all donor animals (n = 6) at 6 dpi. Scale bars are 100 μm .

[Source data](#)

Extended Data Fig. 8 Bodyweight and transmission in hACE2-K18Tg mice.

hACE2-K18Tg mice inoculated with a mixture of wt-S^{614G} and Alpha, or wt-S^{614G} and wt-S^{Alpha}. **(a)** Relative body weight of individual donor mice (n = 4 mice/group; left panel), and contact mice (n = 4 mice/group; right panel). **(b)** Pie chart colors illustrate the ratio of wt-S^{614G} (orange) with Alpha (dark blue), or with wt-S^{Alpha} (light blue) in corresponding experiments in lung homogenates of contact mice at 7 dpc (i.e., 8 dpi of donor mice). Pie chart sizes are proportional to the total viral genome copies reported below. Grey pies indicate values below the LOD (<10³ viral genome copies per sample).

[Source data](#)

Extended Data Fig. 9 Replication of VOC in hACE2-KI mice.

(a–d) Groups hACE2-KI male mice were inoculated intranasally with 10⁴ PFUs of SARS-CoV-2 wt-S^{614G}, Alpha, wt-S^{Alpha} and Beta (n = 8 mice/group). Genome copy numbers in daily oropharyngeal swabs **(a)** and in tissues **(b)**, and virus titers **(c)** in tissues were determined at indicated dpi. Data were log10 transformed and presented as individual values and mean. * p<0.05, **p<0.01 by one-way ANOVA with Tukey's multiple comparisons test comparing the four groups. **(d)** Relative body weight of individual hACE2-KI mice overtime relative to weight at infection (n = 8 mice/group until 2 dpi, and n = 4 mice/group from 3 dpi).

[Source data](#)

Extended Data Fig. 10 Genome sequences of used SARS-CoV-2 variants.

Colors of the variants represent respective viruses in the different experiments. Grey lines indicate positions of known mutations of each virus strain.

Extended Data Table 1 Sequence mutations in S in SARS-CoV-2 recombinant strains

Extended Data Table 2 Sequences of primer and probes for RT-qPCR assays

Extended Data Table 3 Attribution of RT-qPCR assays used for the individual competitive transmission experiments

Supplementary information

Reporting Summary

Peer Review File

Supplementary Table 1

Histopathological score of lungs of hACE2-KI mice infected with single virus inoculum of wt-S^{614G}, Alpha, Beta, or wt-S^{Alpha}. Haematoxilin and eosin-stained slides of left lungs of hACE2-KI mice were scored by a board-certified veterinary pathologist, who was blinded to the identity of the specimen. The scoring scheme is detailed in Supplementary Table 2.

Supplementary Table 2

Lung pathology scoring scheme.

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. 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

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

Ulrich, L., Halwe, N.J., Taddeo, A. *et al.* Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta. *Nature* **602**, 307–313 (2022).
<https://doi.org/10.1038/s41586-021-04342-0>

- Received: 28 June 2021
- Accepted: 13 December 2021
- Published: 22 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04342-0>

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/s41586-021-04342-0>

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

- Article
- Open Access
- [Published: 23 December 2021](#)

Broadly neutralizing antibodies target a haemagglutinin anchor epitope

- [Jenna J. Guthmiller](#) [ORCID: orcid.org/0000-0001-7566-3536](#)¹ na¹,
- [Julianne Han](#) [ORCID: orcid.org/0000-0002-3402-2609](#)² na¹,
- [Henry A. Utset](#)¹,
- [Lei Li](#)¹,
- [Linda Yu-Ling Lan](#)³,
- [Carole Henry](#)¹ nAff16,
- [Christopher T. Stamper](#)³,
- [Meagan McMahon](#)⁴,
- [George O'Dell](#)⁴,
- [Monica L. Fernández-Quintero](#)⁵,
- [Alec W. Freyn](#)⁴ nAff16,
- [Fatima Amanat](#) [ORCID: orcid.org/0000-0002-8029-8227](#)⁴,
- [Olivia Stovicek](#) [ORCID: orcid.org/0000-0001-6413-855X](#)¹,
- [Lauren Gentles](#)^{6,7},
- [Sara T. Richey](#)²,
- [Alba Torrents de la Peña](#)²,
- [Victoria Rosado](#)⁴,
- [Haley L. Dugan](#)³,
- [Nai-Ying Zheng](#)¹,
- [Micah E. Tepora](#)¹,
- [Dalia J. Bitar](#)¹,
- [Siriruk Changrob](#) [ORCID: orcid.org/0000-0002-1826-4333](#)¹,
- [Shirin Strohmeier](#)⁴,
- [Min Huang](#)¹,
- [Adolfo García-Sastre](#) [ORCID: orcid.org/0000-0002-6551-1827](#)^{4,8,9,10,11},
- [Klaus R. Liedl](#) [ORCID: orcid.org/0000-0002-0985-2299](#)⁵,
- [Jesse D. Bloom](#) [ORCID: orcid.org/0000-0003-1267-3408](#)^{6,7,12,13},
- [Raffael Nachbagauer](#) [ORCID: orcid.org/0001-5568-5420](#)⁴ nAff16,

- [Peter Palese](#) ORCID: orcid.org/0000-0002-0337-5823^{4,8},
- [Florian Krammer](#) ORCID: orcid.org/0000-0003-4121-776X⁴,
- [Lynda Coughlan](#) ORCID: orcid.org/0000-0001-9880-6560^{14,15},
- [Andrew B. Ward](#) ORCID: orcid.org/0000-0001-7153-3769² &
- [Patrick C. Wilson](#) ^{1,3} nAff17

Nature volume 602, pages 314–320 (2022)

- 15k Accesses
- 381 Altmetric
- [Metrics details](#)

Subjects

- [Antibodies](#)
- [Immunological memory](#)

Abstract

Broadly neutralizing antibodies that target epitopes of haemagglutinin on the influenza virus have the potential to provide near universal protection against influenza virus infection¹. However, viral mutants that escape broadly neutralizing antibodies have been reported^{2,3}. The identification of broadly neutralizing antibody classes that can neutralize viral escape mutants is critical for universal influenza virus vaccine design. Here we report a distinct class of broadly neutralizing antibodies that target a discrete membrane-proximal anchor epitope of the haemagglutinin stalk domain. Anchor epitope-targeting antibodies are broadly neutralizing across H1 viruses and can cross-react with H2 and H5 viruses that are a pandemic threat. Antibodies that target this anchor epitope utilize a highly restricted repertoire, which encodes two public binding motifs that make extensive contacts with conserved residues in the fusion peptide. Moreover, anchor epitope-targeting B cells are common in the human memory B cell repertoire and were recalled in humans by an oil-in-water adjuvanted chimeric haemagglutinin vaccine^{4,5}, which is a potential universal influenza virus vaccine. To maximize protection against seasonal and pandemic influenza viruses, vaccines should aim to boost this previously untapped source of broadly neutralizing antibodies that are widespread in the human memory B cell pool.

[Download PDF](#)

Main

Antibodies to the major surface glycoprotein haemagglutinin (HA) are critical for providing protection against influenza virus infection^{6,7}. However, most HA-binding antibodies target variable epitopes of the HA head domain, which provide limited protection against antigenically similar influenza virus strains³. Vaccine formulations that preferentially induce antibodies to conserved epitopes of the HA head and stalk domains could provide broad and potent protection against a wide array of influenza viruses. Several leading universal influenza virus candidates are designed to induce antibodies specifically to the stalk domain, but the spectrum of distinct epitopes on the stalk targeted by the human B cell repertoire remains to be determined. By analysing the specificities of B cells targeting the H1 stalk through the generation of monoclonal antibodies (mAbs), our study reveals a new class of broadly neutralizing antibodies (bnAbs) to an underappreciated epitope where HA anchors itself into the viral membrane. Next-generation vaccine platforms should take advantage of this finding to elicit antibodies to the conserved anchor epitope.

Discovery of anchor epitope-binding mAbs

To investigate the specificities of HA-specific antibodies, we generated 358 mAbs from plasmablasts and HA⁺ memory B cells (MBCs) isolated from volunteers who were vaccinated against or naturally infected with seasonal influenza viruses or were participants in a phase I clinical trial of a chimeric HA (cHA) vaccine^{4,5}. Of all mAbs tested, nearly half targeted the HA stalk domain, 21% of which targeted the well-characterized central stalk (CS) epitope (Extended Data Fig. 1a). Notably, stalk-binding mAbs were disproportionately isolated from the infected, 2009 pandemic H1N1 (pH1N1) monovalent inactivated influenza vaccine, and the cHA vaccine cohorts (Extended Data Fig. 1b), as these exposure routes have previously been shown to induce antibody responses to the HA stalk^{5,8,9}. To investigate which epitopes the remaining 79% of stalk-binding mAbs were targeting, we performed negative-stain electron microscopy with three non-CR9114 (ref. 10) competing stalk domain-binding mAbs: 047-09 4F04, 241 IgA 2F04 and 222-1C06. All three mAbs bound an epitope near the anchor of the HA stalk and were oriented at an upward angle towards the epitope (Fig. 1a, b, Extended Data Fig. 1c), suggesting that this epitope may be partially obstructed by the lipid membrane and may only be exposed for antibody binding as the HA trimers flex on the membrane. FISW84, a recently identified stalk-binding mAb¹¹, targets an epitope that overlaps with the three identified anchor-binding mAbs (Extended Data Fig. 1d), suggesting that the anchor epitope is a common stalk epitope. Moreover, a proximal epitope was previously identified on group 2 viruses that is targeted by mAbs CR8020 (ref. 12) and CR8043 (ref. 13). Despite some overlap, the epitope targeted by the group 2 mAbs was considerably

farther up and to the right on the HA stalk relative to the anchor epitope, and mAbs CR8020 and CR8043 targeted the stalk from above (Extended Data Fig. [1e, f](#)), at an angle and positioning more akin to antibodies targeting the CS epitope. mAbs binding the CS epitope (CR9114 (ref. [10](#)) and FI6v3 (ref. [14](#))) did not have overlapping footprints or compete for HA binding with the anchor-binding 047-09 4F04 mAb in an HA competition assay (Fig. [1c, d](#)). In total, we identified 50 distinct mAbs that competed for binding to the anchor epitope from a total of 21 individuals (Fig. [1d](#), Extended Data Tables [1, 2](#)). Of these, 34 anchor-binding mAbs from 15 donors were isolated from the cHA vaccine trial (Extended Data Tables [1, 2](#)).

Fig. 1: The anchor epitope is a common target of stalk-binding antibodies.

 **figure 1**

a, Negative-stain EM of representative 2D class averages and 3D reconstructions of Fabs binding to A/California/7/2009 (E376K) HA. 047-09 4F04 was imaged at $\times 52,000$ normal magnification and 222-1C06 and 241 IgA 2F04 were imaged at $\times 62,000$ normal magnification. **b**, Juxtaposed 3D reconstructions of Fabs binding to

A/California/7/2009 (E376K) HA. **c**, Binding footprints of anchor-binding Fabs relative to mAbs targeting the CS epitope (CR9114 and FI6v3). **d**, Competition of stalk-binding mAbs with CR9114 or 047-09 4F04 (bold mAbs). **e**, Neutralization potency of anchor-binding ($n = 50$) and CS-binding ($n = 37$) mAbs to A/California/7/2009. Data are represented as mean \pm s.d. and were analysed by a two-tailed unpaired non-parametric Mann–Whitney test. IC₅₀, half-maximal inhibitory concentration. **f**, Proportion of anchor-targeting ($n = 50$) and CS-targeting ($n = 37$) mAbs binding to other group 1 influenza virus A subtypes. Data were analysed by Fisher’s exact tests. **g**, Representative 2D class averages ($\times 62,000$ normal magnification), 3D reconstructions and footprints of 222-1C06 binding to H2 and the relative footprint on H1. See also Extended Data Figs. [1–3](#).

[Source data](#)

Anchor-binding mAbs were broadly reactive and broadly neutralizing to pre-pandemic and post-pandemic H1N1 viruses and a swine-origin H1N2 virus (Fig. [1e](#), Extended Data Fig. [1g,h](#)). Notably, anchor-binding mAbs had similar neutralizing potency to pH1N1 relative to mAbs to the CS epitope (Fig. [1e](#), Extended Data Fig. [1i,j](#)). Many stalk-targeting antibodies mediate protection via Fc-mediated functions, including antibody-dependent cellular cytotoxicity^{[15,16](#)}. Anchor epitope-binding mAbs largely did not possess antibody-dependent cellular cytotoxicity activity (15 out of 18; Extended Data Fig. [1k–l](#)), potentially due to the upward angle of approach of anchor-binding antibodies, which may position the Fc distantly from effector cells. Despite pan-H1 binding, anchor-targeting mAbs rarely cross-reacted with other HA subtypes tested, including H3, a group 2 subtype, other group 1 subtypes, including H2 and H5, and influenza B viruses (Fig. [1f](#), Extended Data Fig. [2a](#)). Despite this, the 222-1C06 mAb cross-reacted with H2 and H5 HA (Fig. [1g](#), Extended Data Fig. [2b](#)) and several anchor-binding antibodies could neutralize an H2N2 virus (Extended Data Fig. [2c](#)), suggesting that antibodies targeting the anchor epitope can cross-neutralize other group 1 influenza A viruses. We recently demonstrated that bnAbs to the HA stalk are often polyreactive^{[17](#)}, which may limit the activation of B cells expressing bnAbs. Relative to mAbs that target the CS epitope, we identified that anchor-binding mAbs were proportionally less likely to be polyreactive and those that were polyreactive had weaker relative affinity for lipopolysaccharide (Extended Data Fig. [2d,e](#)). These data suggest that, although polyreactivity is selected for within the anti-anchor epitope B cell pool, it is not to the same level as B cells to the CS epitope.

H1N1 viruses have acquired several mutations within the HA stalk domain, probably due to antibody selective pressures or to increase stability. To understand whether these mutations affect antibody binding to the anchor epitope, we screened mAbs against naturally occurring and experimentally identified viral escape mutants of mAbs binding to the CS epitope (Extended Data Fig. [2f,g](#), Extended Data Table [3](#)).

Anchor epitope-binding mAbs were mostly unaffected by these mutants, whereas most of the CS-binding mAbs showed reduced binding to at least one mutant (Extended Data Fig. 2g). Notably, most mAbs had reduced binding to A373V of HA2, which has recently been shown to preferentially grow in the presence of mAbs to the CS epitope². While A373 is distant from the anchor epitope, the A373V mutation has been shown to affect the conformation of the HA stalk², explaining the broad reduction of HA binding by antibodies targeting either stalk epitope. Anchor-binding mAbs only demonstrated a 10–30% reduction in binding (Extended Data Fig. 2g), indicating that they are still likely to neutralize viruses carrying the A373V mutation.

To test whether mAbs targeting the anchor epitope were protective in vivo, we administered a cocktail of five mAbs targeting the anchor epitope or the CS epitope prophylactically and therapeutically (48 h after infection) to mimic a polyclonal response against these epitopes and infected mice with a lethal dose of a mouse-adapted pH1N1 virus (A/Netherlands/602/2009; Supplementary Table 1). Mice that received a prophylactic or therapeutic anti-anchor cocktail at 5 mg/kg experienced 100% protection from weight loss and lethal infection (Extended Data Fig. 3a, b). No differences in lung viral titres were detected in mice that received the anti-anchor cocktail prophylactically relative to mice administered the negative control mAb (Extended Data Fig. 3c). Anti-stalk antibodies do not provide sterilizing immunity but are known to neutralize subsequent rounds and limit disease progression¹⁸. As a result, lung viral titres do not necessarily reflect protection from morbidity and mortality. Finally, the anti-anchor cocktail given prophylactically provided 70% protection against lethal A/Fort Monmouth/1/1947 infection (Extended Data Fig. 3d), a virus that circulated before the birth of any of the donors in our study (Extended Data Table 1). Since anchor-binding mAbs do not engage in antibody-dependent cellular cytotoxicity for the most part, antibodies that target the anchor epitope probably provide protection in vivo through direct neutralization of the virus. Together, these data indicate that antibodies to the anchor epitope are pan-H1 neutralizing and protective in vivo.

Structure of an anchor-binding antibody

All anchor epitope-binding mAbs utilized one of four VH3 genes: VH3-23, VH3-30 or VH3-30-3, and VH3-48, in contrast to mAbs targeting the CS epitope, which commonly use VH1-69 (Fig. 2a, Extended Data Fig. 4a, b). Anchor epitope-binding mAbs also utilized a highly restricted light chain repertoire, with all mAbs utilizing a combination of VK3-11 or VK3-15 combined with JK4 or JK5 (Fig. 2b, Extended Data Fig. 4c, d). Furthermore, all but one light chain of the anchor-targeting mAbs were clonally related (Extended Data Table 2), indicating that the light chains were very similar across mAbs and study participants. We identified four distinct clonal expansions, with one public clone found across two donors (Extended Data Fig. 4e, f, Extended Data Table 2). Anchor epitope-targeting and CS-targeting mAbs exhibited

similar levels of somatic hypermutations (Extended Data Fig. 4g). The κ -chain complementarity-determining region 3 (K-CDR3) length of anchor epitope-binding mAbs was highly restricted, with all K-CDR3s being ten amino acids long (Extended Data Fig. 4h).

Fig. 2: Anchor-targeting mAbs bind to the HA fusion peptide via public binding motifs.

 figure 2



a, b, VH (**a**) and VK (**b**) gene usage by anchor-binding mAbs. The number in the centre of the pie graphs indicates the number of mAbs analysed. **c**, Cryo-EM structure of anchor-targeting 222-1C06 (blue) and lateral patch-targeting 045-09 2B05 (dark grey; see [Methods](#)) binding to A/California/7/2009 (E376K) HA (light grey). **d**, Heavy chain and light chain footprint of 222-1C06 binding to HA based on the cryo-EM structure. **e**, HA epitope contact residues (maroon) and heavy chain (green) and light chain (yellow) antibody contact residues of the 222-1C06 paratope. Peach-highlighted amino acids represent the fusion peptide of HA2. **f**, K-CDR3 NWP and H-CDR2 Y58 motifs of 222-1C06. Bold residues are HA residues. **g**, Major contacts of 222-1C06 K-CDR1 and H-CDR3 (normal typeface) binding to HA (bold residues). **h, i**, Weblogo plot and germline sequence of Y58 following the H-CDR2 motif (**h**) and the K-CDR3

NWP motif (**i**). **j, k**, Amino acid conservation of s contact residues across human, swine and avian H1 viruses (**j**) and group 1 influenza A viruses (**k**). Bold residues are contacts conserved with A/California/7/2009 H1N1 (**k**). The strain information used for conservation models in **j** and **k** are in Supplementary Tables [4](#), [5](#), respectively. See also Extended Data Fig. [4](#).

Source data

To investigate the binding motif of anchor-targeting mAbs, we generated a high-resolution (3.38 Å) cryo-electron microscopy (cryo-EM) structure of 222-1C06 bound to A/California/7/2009 HA (Fig. [2c](#), Extended Data Fig. [4i–k](#)). The broad paratope of 222-1C06 made extensive contacts across the HA fusion peptide¹⁹ (Fig. [2d, e](#)), which mediates viral membrane fusion with the host membrane. Binding of the fusion peptide was largely mediated by an NWP motif within the K-CDR3, a Y58 directly following the H-CDR2, and a W99 in the H-CDR3, with these HA-binding motifs acting independently and in combination via an aromatic pocket (Fig. [2f, g](#), Supplementary Table [2](#)). Moreover, both the K-CDR3 NWP and the H-CDR2 Y58 were found in all the anchor-binding mAbs and were germline encoded (Fig. [2h, i](#)), which could have led to the selection of B cells utilizing these variable genes. Notably, FISW84 utilizes VH3-23 and VK3-15, and comparison of the paratopes showed that the NWP and Y58 motifs of FISW84 similarly form an aromatic pocket but orient towards the fusion peptide slightly differently than 222-1C06 (Extended Data Fig. [4l](#)). Molecular dynamics simulations also showed that VH3-23-utilizing and VH3-30-utilizing mAbs from our study and FISW84 targeted the HA fusion peptide similarly to the cryo-EM structure of 222-1C06 via the aromatic pocket created by the K-CDR3 NWP and H-CDR2 Y58 motifs, albeit at different orientations (Extended Data Fig. [4m](#)). Crucially, molecular dynamic and cryo-EM analyses revealed numerous intra-Fab interactions of hydrophobic and aromatic amino acids, including p-stacking of K-CDR3-P95 with K-CDR3-W94 and H-CDR2-Y58 that rigidified the paratope (Extended Data Fig. [4n, o](#), Supplementary Table [3](#)).

Broad analysis of human, swine and avian H1 viruses revealed that the side-chain contacts of 222-1C06 were highly conserved (94–100% conserved; Fig. [2j](#)). In addition, the side-chain contacts were 100% conserved across 100 years of H1N1 virus evolution in humans (Extended Data Fig. [4p](#)). Deep mutational scanning of potential H1 viruses at these contact residues indicated substantial permissibility (Extended Data Fig. [4q](#)), although these mutations appear to not have been selected for in nature (Fig. [2j](#)). The five side-chain contacts of this broad epitope were also highly conserved across group 1 viruses, with the W343 contact being 100% conserved across all group 1 viruses (Fig. [2k](#)). Together, these data reveal that B cells targeting the anchor epitope utilized a highly restricted V(D)J gene repertoire, and the specific features within this repertoire made critical and extensive contacts with the conserved anchor epitope.

The anchor epitope is a common target

Owing to the restricted repertoire features of anchor-targeting mAbs, we next determined the relative proportion of B cells with these features by interrogating single-cell repertoire sequencing of HA-specific B cells isolated from 20 human participants following cHA vaccination^{4,5}. The cHA vaccine platform is intended to specifically induce antibodies to the stalk domain by retaining the stalk domain of H1 and replacing the head domain of HA with that of an avian subtype, either H8 (prime) or H5 (boost) for this trial^{4,5}. To investigate the proportion of B cells with anchor epitope-binding repertoire features, we selected B cells that used VH3-23/VH3-30/VH3-30-3/VH3-48, VK3-11/VK3-15, JK4/JK5, a K-CDR3 ten amino acids in length, and possessed an NWP motif within the K-CDR3. We also segregated B cells expressing VH1-69 and a κ-chain, which are commonly used by B cells that target the CS epitope. We identified that B cells with features of antibodies binding the anchor epitope were abundant within the human B cell repertoire, with 6% of all B cells identified fitting within this defined repertoire (Fig. 3a). Moreover, all but one participant had at least one B cell with anchor-binding repertoire features (Fig. 3b). In addition, 32 out of 33 mAbs generated from the selected anchor-targeting B cell list competed for HA binding with 047-09 4F04 (Extended Data Fig. 5a), indicating that this population was greatly enriched for anchor-targeting B cells. Moreover, the anchor-binding B cells were highly mutated and were largely class-switched to IgG1 (Extended Data Fig. 5b,c), suggesting that these B cells were MBCs that had undergone affinity maturation and class-switch recombination. Together, these data indicate that the anchor epitope is a common target of the human MBC repertoire following cHA vaccination.

Fig. 3: MBCs and serum antibodies commonly target the anchor epitope.

 **figure 3**

a, Proportion of all cH5/1⁺ B cells with repertoire features of anchor-binding mAbs, VH1-69/κ (CS epitope) or with other repertoire features. The number in the centre of the pie graph is the number of B cells analysed. **b**, Proportion of B cells with anchor-binding mAb features or that use VH1-69/κ-chain by participant ($n = 20$ donors). Lines connect the same participant. Data were analysed using a two-tailed paired non-

parametric Wilcoxon matched-pairs signed rank test. **c**, Electron microscopy polyclonal epitope mapping (EMPEM) summary of polyclonal antibodies (pAbs) binding to A/Michigan/45/2015 HA in the serum of participants 236 and 241 collected at day 7 and day 14 following 2014 quadrivalent inactivated influenza vaccine. Fabs shown as graphics with dotted lines represent predicted placements due to limited particle representation. 2D class averages were imaged at $\times 62,000$ normal magnification **d**, Overlap of 241 IgA 2F04 Fab and pAb binding the anchor epitope from participant 241. See also Extended Data Fig. 5.

Source data

To confirm that anchor epitope-targeting mAbs were representative of the serum antibody response, we performed electron microscopy polyclonal epitope mapping (EMPEM)²⁰ with serum antibodies from participants 236 and 241 from the 2014 quadrivalent inactivated influenza vaccine cohort (Extended Data Table 1). Both participants had detectable antibodies targeting the anchor epitope at days 7 and 14 post-vaccination (Fig. 3c, Extended Data Fig. 5d,e). Comparison of anchor epitope-binding polyclonal antibodies identified in participant 241 revealed that the 241 IgA 2F04 mAb strongly overlapped with the 241 polyclonal antibody (Fig. 3d), suggesting that the polyclonal antibody derived from this clonal expansion. Together, these data indicate that humoral immunity against the anchor epitope is common within the MBC pool and polyclonal serum antibody response after vaccination.

cHA induces anchor-binding antibodies

To investigate whether participants enrolled in a phase I clinical trial of the cHA vaccine (Fig. 4a) mounted an antibody response to the anchor and the CS epitope, we adapted the competition ELISA to detect serum antibody responses that could compete for binding with 047-09 4F04 and CR9114, respectively. Three different vaccine formulations were used in this trial, with participants being primed with either a cH8/1 inactivated influenza vaccine (IIV) with an oil-in-water adjuvant (AS03) or a cH8/1 live-attenuated influenza vaccine followed by a boost with a cH5/1 IIV with or without AS03 (Fig. 4a). Only participants who received the IIV + AS03 on either the prime or boost were capable of seroconverting against both the anchor and the CS epitopes (Fig. 4b,c, Extended Data Fig. 6a,b). Participants who received the cH8/1 IIV + AS03 prime did not further increase serum antibodies to either stalk epitope after the cH5/1 + AS03 boost (Fig. 4c), suggesting that these B cells were refractory to continued activation. Serum titres against the anchor and CS epitopes (Extended Data Fig. 6a,b) closely matched that of serum neutralizing titres against a cH6/1N8 virus and an avian-swine H1N1 virus⁵, suggesting that the anchor-targeting and CS-targeting serum antibodies were responsible for neutralization.

Fig. 4: cHA vaccination in humans robustly recalls MBCs targeting the anchor epitope.

 figure 4

a, cHA vaccine trial design including vaccine groups (right; group 1 $n = 10$ participants; group 2 $n = 7$ participants; group 4 $n = 7$ participants; group 3 and 5 $n = 6$). i.m., intramuscular; i.n., intranasal; LAIV, live-attenuated influenza vaccine; PBS, phosphate buffered saline. Bottle images created with BioRender.com. **b, c**, Fold change by participant of serum antibodies competing for binding to the anchor and CS epitopes after the prime (d29/d1; **b**) and the boost (d113/85; **c**). **d, e**, Proportion of participants who seroconverted (**d**) and half-maximal effective concentration titres (EC₅₀; **e**) to the anchor and CS epitopes. Individuals in the cHA (IIV + AS03) cohort were those who received the cHA vaccine with adjuvant (cH8/1 IIV + AS03 prime and cH5/1 IIV + AS03 boost; $n = 17$ donors) and the IIV cohort were those who received licensed IIV vaccines (2009 monovalent inactivated influenza vaccine, 2010 trivalent inactivated influenza vaccine and 2014 quadrivalent inactivated influenza vaccine; $n = 11$ donors). Data in **b, c** and **e** are mean \pm s.d. The dotted line represents the limit of detection. Data in **b** and **c** were analysed by two-tailed non-parametric Kruskal–Wallis tests. Data in **d** were analysed by Fisher’s exact test. Data in **e** were analysed by two-tailed unpaired non-parametric Mann–Whitney test. See also Extended Data Fig. [6](#).

Source data

All but one participant in the IIV + AS03 groups seroconverted against the anchor epitope (Fig. 4d) and had higher titres against the anchor epitope than participants who received the 2009 monovalent inactivated influenza vaccine or seasonal influenza virus vaccines (Fig. 4e). However, the precise role of the cHA immunogen versus the AS03 adjuvant in inducing anti-stalk antibody responses could not be resolved in our study. Notably, those individuals who received the IIV alone had weak plasmablast responses relative to those individuals who received the IIV + AS03 (ref. 4), suggesting that the oil-in-water adjuvant, not the cHA immunogen, was essential for robust activation of B cells and anti-anchor antibody responses. Moreover, considerable pre-existing antibodies may hinder recall of B cells to the stalk. Individuals first exposed to the 2009 pH1N1, a virus for which individuals had low pre-existing humoral immunity, had proportionally more isolated mAbs to the stalk and were more likely to have an anti-anchor mAb than individuals who had repeatedly been exposed to the pH1N1 virus in subsequent influenza seasons (Extended Data Fig. 6c,d), therefore suggesting that pre-existing antibodies may limit antibody responses to the HA stalk. Despite robust recall of antibodies to the anchor and CS epitopes by the adjuvanted vaccines, titres decreased 1 year after vaccination (day 420; Extended Data Fig. 6e,f). Together, these data indicate that the adjuvanted inactivated cHA vaccine can robustly induce antibodies to multiple stalk epitopes, including the anchor.

Headless HA antigens, including mini-HA²¹, are attractive universal influenza virus vaccine antigens, as these antigens lack the immunodominant epitopes of the HA head^{21,22}. However, only 1 out of 50 anchor-binding mAbs bound the mini-HA antigen²¹, whereas all anchor epitope-binding mAbs bound cH6/1 (Extended Data Fig. 6g). Compared to full-length HA, the membrane-proximal region of the mini-HA splays by approximately 14.5 Å (ref. 21), which may disrupt the antigenicity of the anchor epitope. To understand whether anchor epitope-targeting antibodies could bind to the mini-HA in a more native setting, we generated a membrane-bound mini-HA and observed that mAbs targeting the anchor and CS epitopes readily bind to both the full-length membrane-bound A/California/7/2009 HA and the membrane-bound mini-HA (Extended Data Fig. 6h), indicating that the mini-HA is antigenic when natively presented. Furthermore, we demonstrated that anchor epitope-targeting antibodies bound HA with a fibritin but not a GCN4 trimerization domain (Extended Data Fig. 6i), highlighting selection of the trimerization domain as an important factor for vaccine design. Together, these data demonstrate that native-like HA antigens, such as the cHA vaccine, can recall MBCs that target the anchor epitope.

Discussion

In this study, we identified a public class of bnAbs that target an epitope at the anchor of the HA stalk domain near the membrane. The anchor-targeting mAbs were public clonotypes across participants, with all antibodies possessing two conserved, germline-encoded binding motifs: a Y58 directly following the H-CDR2 and an NWP motif within the K-CDR3. The neutralizing activity of anchor epitope-targeting mAbs to pre-pH1N1 and post-pH1N1 viruses and a swine-origin H1-expressing virus indicates that the anchor epitope is an important target for pan-subtype neutralizing antibodies. As two of the last four influenza virus pandemics were caused by H1N1 viruses and a recent report has shown that antigenically novel H1-expressing viruses commonly spill over from swine into humans²³, it is critically important to generate pan-H1 vaccines to prevent the next influenza pandemic. Moreover, the ability of anti-anchor antibodies to neutralize an H2 virus and the general conservation of contact residues suggests that anchor-targeting antibodies have the potential to acquire cross-neutralization against group 1 viruses.

Our study highlights an additional broadly protective epitope of the HA stalk and provides guidance on how vaccines can be designed to drive broadly protective antibodies to multiple distinct epitopes, which can work cooperatively to provide optimal protection while avoiding the generation of antibody escape mutants. Notably, studies in the HIV-1 field have shown that bnAb monotherapy can lead to the development of antibody-resistant viral variants^{24,25,26}, whereas combination bnAb therapy does not²⁷. In addition, immune focusing towards a single epitope can lead to the generation and selection of viral escape mutants at these highly conserved epitopes^{23,28,29}. Therefore, it is critical that future universal influenza virus vaccines elicit antibodies to multiple conserved epitopes to prevent the generation of bnAb escape viruses.

The angle of approach of anchor-binding mAbs and the proximity of the epitope to the viral membrane suggest that this epitope is typically obstructed, limiting antibody recognition and B cell activation. However, membrane-bound HA typically flexes between 0° and 25° and up to 52° on its threefold axis¹¹, suggesting antibodies and B cells can easily access the anchor epitope during this flexing process. Moreover, H1 viruses possess a highly conserved glycosylation site on the HA stalk that lies above the anchor epitope^{11,30}, which may obstruct antibodies from targeting this epitope from above. Similarly, a neutralizing mAb to the Middle East respiratory syndrome coronavirus targets an epitope on S2 from an upward angle to avoid glycans³¹. Therefore, the upward angle of approach may be a common feature of antibodies that recognize epitopes below glycans.

Our study shows that humans have pre-existing immunity against the anchor epitope and influenza virus vaccination can recall MBCs to secrete antibodies to this epitope. However, vaccine HA antigens must have a native-like conformation near the

transmembrane domain, as trimer splaying potentially due to the GCN4 trimerization domain ablates antibody binding to the anchor epitope. Moreover, our study reveals that the cHA vaccine strategy recalled MBCs to the anchor and CS epitopes, as these MBCs do not need to compete against MBCs that target the immunodominant variable HA head epitopes^{3,9,17,32}. Similarly, the mini-HA/headless HA vaccine strategy has the potential to also recall MBCs to multiple epitopes of the HA stalk domain, if displayed natively^{22,33}. The addition of an oil-in-water AS03 adjuvant to the cHA was critical for recalling MBCs to the anchor epitope. Oil-in-water adjuvants largely function to emulsify antigen, which may prevent sequestration of antigen by circulating antibodies, increase delivery of free antigen to lymph nodes, and help to stimulate innate immune receptors^{34,35}. Notably, an oil-in-water adjuvanted inactivated H5N1 vaccine induced higher neutralizing antibody titres, antibodies to more HA epitopes, and higher avidity antibodies³⁶. Therefore, the inclusion of oil-in-water adjuvants may have an important role in generating bnAbs and may improve vaccine effectiveness of seasonal influenza vaccines. Together, our study shows that influenza vaccination strategies, such as the cHA vaccine with the AS03 adjuvant, have the capability to robustly induce antibodies to the previously underappreciated anchor epitope and can provide broad protection against H1 viruses.

Methods

Study approvals, cohorts and human materials

Human peripheral blood mononuclear cells (PBMCs) and serum were obtained from multiple donors from multiple cohorts, which is outlined in Extended Data Table 1. Informed consent was obtained from all participants. All studies were performed with the approval of the University of Chicago Institutional Review Board (IRB; ID #09-043-A). The 2009 pH1N1 infection and 2009 monovalent inactivated influenza vaccine (MIV) cohorts were also approved by the IRB at Emory University. The chimeric HA vaccine study cohort is identified as clinical trial NCT03300050 and further details on trial design are outlined elsewhere^{4,5}. The study was approved by IRBs at local clinical sites, including Icahn School of Medicine at Mount Sinai, Duke University, and Cincinnati Children's Hospital Medical Center. All experiments performed with mice were done in accordance with the University of Chicago and Icahn School of Medicine at Mount Sinai Institutional Animal Care and Use Committees.

Cell lines

Human embryonic kidney HEK293T (female, #CRL-11268), Madin Darby canine kidney (MDCK; female, #CCL-34, NBL-2) and human A549 (#CCL-185) cells were

purchased and authenticated by the American Type Culture Collection (ATCC). MDCK-SIAT1 cells were generated previously³⁷. All cells were maintained in a humidified atmosphere of 5% CO₂ at 37 °C. HEK293T cells were maintained in advanced-DMEM supplemented with 2% ultra-low IgG fetal bovine serum (FBS; Invitrogen), 1% l-glutamine (Invitrogen) and 1% antibiotic-antimycotic (Invitrogen). MDCK, MDCK-SIAT1 and A549 cells were maintained in DMEM supplemented with 10% FBS (Invitrogen), 1% l-glutamine (Invitrogen) and 1% penicillin–streptomycin (Invitrogen). Jurkat cells expressing FcgRIIIa and FcgRI with a NFAT-driven luciferase reporter gene (#G7010) were acquired and validated by Promega and were directly used for antibody-dependent cellular cytotoxicity (ADCC) assays. Cell lines were not authenticated after receiving from suppliers and were not tested for mycoplasma.

Monoclonal antibody production

Monoclonal antibodies were generated as previously described^{38,39,40}. Peripheral blood was obtained from each donor approximately 7 days after vaccination or infection or obtained 28⁺ days post-vaccination. Lymphocytes were isolated and enriched for B cells using RosetteSep. Total plasmablasts (CD3⁻CD19⁺CD27^{hi}CD38^{hi}; all cohorts except 2014 quadrivalent inactivated influenza vaccine (QIV)), IgG⁺ plasmablasts (CD3⁻CD19⁺IgM⁻CD27^{hi}CD38^{hi}IgG⁺IgA⁻; 2014 QIV), IgA⁺ plasmablasts (CD3⁻CD19⁺IgM⁻CD27^{hi}CD38^{hi}IgG⁻IgA⁺; 2014 QIV cohort), or HA⁺ bait-sorted MBCs (CD3⁻CD19⁺CD27⁺CD38^{lo/+}HA⁺; for 030-09M 1B06) were single-cell sorted into 96-well plates. Genes encoding immunoglobulin heavy and light chains were amplified by reverse transcriptase PCR (RT-PCR), sequenced, cloned into human IgG1, human κ-chain, or human λ expression vectors, and co-transfected into HEK293T cells. Secreted mAbs were purified from the supernatant using protein A agarose beads. For mAbs generated from the 2014 QIV cohort, mAb names include the original isotype of the sorted plasmablast, and all mAbs were expressed as human IgG1. cH5/1-binding B cells (CD19⁺CD27⁺cH5/1⁺) were sorted from donors 28 days after cH5/1 vaccination (NCT03300050). Cells were sorted with A/California/04/2009 HA probe (for 030-09M 1B06) or cH5/1 probe, each with a Y98F mutation to ablate non-specific binding to sialic acids on B cells. mAb heavy chain and light chain sequences were synthesized from single-cell RNA sequencing data of cH5/1-baited B cells (IDT), and cloned into the human IgG1, human κ-chain or human λ expression vectors. B cell clones were determined by aligning all the V(D)J sequences sharing identical progenitor sequences, as predicted by IgBLAST using our in-house software, VGenes. Consensus sequence analysis was performed using WebLogo⁴¹ and sequence alignments were determined using Clustal Omega.

Viruses and recombinant proteins

Influenza viruses used in all assays were grown in-house in specific pathogen free (SPF) eggs, harvested, purified and titred. Recombinant HA, cHA and mini-HA were obtained from BEI Resources or generated in-house. Recombinant HA mutant proteins used in Extended Data Fig. 2 were generated with identified mutations from the deep mutational scanning experiments (see below) or with known mutations that have arisen naturally or were identified in other studies^{2,10,42,43,44,45,46,47,48,49,50,51,52} (Extended Data Table 3). All mutations were made on HA from A/California/7/2009 and were expressed in HEK293T cells and purified using Ni-NTA agarose beads (Qiagen).

Antigen-specific ELISA

High protein-binding microtitre plates (Costar) were coated with 8 haemagglutination units (HAU) of virus in carbonate buffer or with recombinant HA, including HA mutants described below, at $2 \mu\text{g ml}^{-1}$ in phosphate-buffered saline (PBS) overnight at 4°C . Plates were washed the next morning with PBS 0.05% Tween and blocked with PBS containing 20% FBS for 1 h at 37°C . Antibodies were then serially diluted 1:3 starting at $10 \mu\text{g ml}^{-1}$ and incubated for 1.5 h at 37°C . Horseradish peroxidase (HRP)-conjugated goat anti-human IgG antibody diluted 1:1,000 (Jackson Immuno Research) was used to detect binding of mAbs, and plates were subsequently developed with Super Aquablue ELISA substrate (eBiosciences). Absorbance was measured at 405 nm on a microplate spectrophotometer (Bio-Rad). To standardize the assays, control antibodies with known binding characteristics were included on each plate, and the plates were developed when the absorbance of the control reached 3.0 optical density (OD) units. All ELISAs were performed in duplicate twice. To define antibodies as targeting the H1 stalk, mAbs were tested for binding to cH5/1, which utilizes the head domain from H5-expressing viruses and the stalk domain from the pH1N1 virus⁵³, and for haemagglutination inhibition (HAI) activity against pH1N1 (A/California/7/2009). mAbs that bound the cHA and lacked HAI activity were classified as those binding the HA stalk domain. To classify antigen specificity, mAbs that did not definitively bind to the HA head or stalk are listed as binding unknown HA⁺ epitopes. Affinity measurements, as represented as dissociation constant (K_d) at a molar concentration (M), were calculated using Prism 9 (GraphPad) by performing a non-linear regression. All experiments were performed in duplicate and technically replicated twice.

Competition ELISAs

Plates were coated with $50 \mu\text{l}$ of A/California/7/2009 HA at a concentration of $1 \mu\text{g ml}^{-1}$ and incubated overnight at 4°C . To biotinylate the antibodies with known

epitope specificities, CR9114 (CS epitope) and 047-09-4F04 (anchor epitope) were incubated at 4 °C with EZ-Link Sulfo-NHS-Biotin (Thermo Scientific) for 24 h or 48 h before use, respectively. After blocking the plates with PBS containing 20% FBS for 1 h at 37 °C, serum samples were incubated (starting dilution of 1:50 for human serum or 20 µg ml⁻¹ for mAbs) in the coated wells for 2 h at room temperature. Either biotinylated CR9114 or 047-09 4F04 was then added at a concentration equal to twice its K_d and incubated in the wells with the serum or mAbs for 2 h at room temperature. The biotin-ylated antibodies were desalting before addition to remove free biotin using Zeba spin desalting columns, 7 k MWCO (Thermo Scientific). After washing the plates, wells were incubated with HRP-conjugated streptavidin (Southern Biotech) at 37 °C for 1 h for detection of the biotin-ylated antibody. Super Aquablue ELISA substrate (eBiosciences) was then added, and absorbance was measured at 405 nm on a microplate spectrophotometer (Bio-Rad). To standardize the assays, biotinylated CR9114 or 047-09 4F04 was incubated in designated wells on each plate without any competing serum or mAb, and data were recorded when the absorbance of these wells reached an OD of 1–1.5 units. After subtracting background, percent competition by serum samples was then determined by dividing the observed OD of a sample by the OD reached by the positive control, subtracting this value from 1, and multiplying by 100. For the serum data, ODs were log transformed and analysed by non-linear regression to determine EC₅₀ values using Prism software (GraphPad). For Fig. 4 and Extended Data Fig. 5, only donors with serum for all timepoints were included. All experiments were performed in duplicate and technically replicated twice.

Polyreactive ELISAs

High-protein binding microtitre plates (Costar) were coated with 10 µg ml⁻¹ calf thymus double-stranded DNA (dsDNA; Thermo Fisher Scientific), 2 µg ml⁻¹ *Salmonella enterica* serovar Typhimurium flagellin (Invitrogen), 5 µg ml⁻¹ human insulin (Sigma-Aldrich), 10 µg ml⁻¹ KLH (Invitrogen) and 10 µg ml⁻¹ *Escherichia coli* LPS (Sigma-Aldrich) in PBS. Plates were coated with 10 µg ml⁻¹ cardiolipin in 100% ethanol and allowed to dry overnight. Plates were washed with water and blocked with PBS, 0.05% Tween and 1 mM EDTA. mAbs were diluted 1 µg ml⁻¹ in PBS and serially diluted fourfold and added to plates for 1.5 h. Plates were washed and goat anti-human IgG-HRP (Jackson Immunoresearch) was diluted 1:2,000 in PBS, 0.05% Tween and 1 mM EDTA. Plates were washed with water and were blocked with PBS, 0.05% Tween and 1 mM EDTA for 5 min. Plates were washed again with water and were developed with Super Aquablue ELISA substrate (eBioscience) until the positive control mAb, 3H9 (ref. 54), reached an A_{450} of 3. All experiments were performed in duplicate and technically replicated twice.

Deep mutational scanning for stalk domain mutants

The mutant libraries used herein were previously described⁵⁵. The libraries consist of all single amino acid mutations to A/WSN/1933 (H1N1). The experiments were performed by using biological triplicate libraries. The mutational antigenic profiling of 045-09 2B06, a CS epitope-binding mAb, was performed as previously outlined⁵⁶. In brief, 10^6 tissue culture infectious dose 50 (TCID₅₀) of two of the virus library biological replicates was diluted in 1 ml in IGM (Opti-MEM supplemented with 0.01% FBS, 0.3% BSA and 100 mg ml⁻¹ calcium chloride) and incubated with an equal volume of 045-09 2B06 antibody at a final concentration of 50 or 25 µg ml⁻¹ for 1.5 h at 37 °C. MDCK-SIAT1 cells were infected with the virus antibody mixtures. Two hours post-infection, the media were removed, the cells washed with 1 ml PBS, and 2 ml of fresh IGM was added. Fifteen hours post-infection, viral RNA was extracted, reverse-transcribed using primers WSNHA-For (5'-AGCAAAAGCAGGGAAAATAAAACAAAC-3') and WSNHA-Rev (5'-AGTAGAAACAAGGGTGTTCCTTATATTCTG-3'), and PCR amplified according to the barcoded-subamplicon library preparation as previously described⁵⁵. The overall fraction of virions that survive antibody neutralization was estimated using quantitative RT-PCR (qRT-PCR) targeting the viral nucleoprotein (NP) and cellular GAPDH as previously described⁵⁶. Using tenfold serial dilutions of the virus libraries, we infected cells with no antibody selection to serve as a standard curve of infectivity. qPCR Ct values from the standard curve samples compared to the virus-antibody mix samples were determined for NP and GAPDH. We then generated a linear regression to fit the difference between the NP and GAPDH Ct values for the standard curve samples, and then used this curve to interpolate the fraction surviving for the antibody-virus selection samples. Across the three library replicates, the fraction of virus surviving antibody selection was 0.17, 0.1 and 0.14.

Illumina(R) deep sequencing data were analysed using dms_tools2 version 2.4.12 software package⁵⁷, which can be found at https://github.com/jbloomlab/dms_tools2. The computer code used is at https://github.com/jbloomlab/2B06_DMS, and the Jupyter notebook that performed most of the analysis is at https://github.com/jbloomlab/2B06_DMS/blob/master/analysis_notebook.ipynb. The sequencing counts were processed to estimate the differential selection for each mutation, which is the log enrichment of that mutation in the antibody-selected condition versus the control⁵⁶. The numerical measurements of the differential selection that 2B06 imposes on each mutation can be found at: https://github.com/jbloomlab/2B06_DMS/blob/master/results/diffsel/tidy_diffsel.csv.

Deep mutational scanning for H1 variants

Amino acid preferences for the HA of A/WSN/1933 (H1N1) were previously determined⁵⁵. In brief, deep mutational scanning was performed by passaging virus

libraries at a low multiplicity of infection in MDCK-SIAT1 cells. Following deep sequencing of the resulting virus, amino acid preferences were determined using the Python package dms_tools (http://jbloomlab.github.io/dms_tools/), version 1.1.12. This program aligns subamplicon reads to the reference HA sequence, counts the number of mutations at each amino acid site, and determines amino acid preferences based on the mutation counts pre-selection and post-selection.

Microneutralization assays

Microneutralization assays for mAb characterization were carried out as previously described^{58,59}. MDCK cells were maintained in DMEM supplemented with 10% FBS, 1% penicillin–streptomycin, and 1% l-glutamine at 37 °C with 5% CO₂. The day before the experiment, 25,000 MDCK cells were added to each well of a 96-well plate. Serial twofold dilutions of mAb were mixed with an equal volume of 100 TCID₅₀ of virus for 1 h and added to MDCK cells for 1 h at 37 °C. The mixture was removed, and cells were cultured for 20 h at 37 °C with 1X MEM supplemented with 1 µg ml⁻¹ tosyl phenylalanyl chloromethyl ketone (TPCK)-treated trypsin and appropriate mAb concentration. Cells were washed twice with PBS, fixed with 80% ice-cold acetone at –20 °C for at least 1 h, washed three times with PBS, blocked for 30 min with 3% BSA–PBS, and then treated for 30 min with 2% H₂O₂. Cells were incubated with a mouse anti-NP antibody (1:1,000; Millipore) in 3% BSA–PBS for 1 h at room temperature, followed by goat anti-mouse IgG HRP (1:1,000; Southern Biotech) in 3% BSA–PBS for 1 h at room temperature. The plates were developed with Super Aquablue ELISA substrate at 405 nm until virus-only controls reached an OD of 1. The signal from uninfected wells was averaged to represent 100% inhibition. The signal from infected wells without mAb was averaged to represent 0% inhibition. Duplication wells were used to calculate the mean and s.d. of neutralization, and the half-maximal inhibitory concentration (IC₅₀) was determined by a sigmoidal dose–response curve. The inhibition ratio (%) was calculated as: ((OD positive control – OD sample)/(OD positive control – OD negative control)) × 100. The final IC₅₀ was determined using Prism software (GraphPad). All experiments were performed in duplicate and technically replicated twice.

H2N2 neutralization assays

Twenty thousand MDCK cells were seeded per well in a 96-well cell culture plate (Corning) and the cells were used the next morning to perform the neutralization assay. Antibody dilutions were prepared starting at 30 µg ml⁻¹ with threefold subsequent dilutions in 1X MEM. Each respective dilution was mixed with 10,000 plaque-forming units (PFU) of cold-adapted A/Ann Arbor/6/1960 (H2N2) virus for 1 h at room temperature. After 1 h, cells were washed with PBS and 100 µl of antibody–

virus mixture was added onto the cells for 1 h at 37 °C. Next, the antibody–virus mixture was removed and 60 µl of 1X MEM containing TPCK was added to each well. Of each corresponding antibody dilution, 60 µl was also added to each well and the cells were incubated at 33 °C for 3 days. On the third day, a haemagglutination assay was performed using turkey red blood cells to assess the HAU at each antibody dilution.

In vivo challenge infections

mAb cocktails (Extended Data Fig. 2b) were passively transferred into 6–8-week-old female BALB/c mice (Jackson Laboratories) by intraperitoneal injection of 0.2, 1 and 5 mg per kg mAb cocktail, which are further detailed in Supplementary Table 1. Negative control mice received 5 mg per kg of the anthrax-specific mAb 003-15D03 as an isotype control. mAbs were administered 2 h before infection for prophylactic treatment and 48 h post-infection for therapeutic treatment. For prophylactic mAb studies with A/Netherlands/602/2009 (Extended Data Fig. 3a), mice were anaesthetized with isoflurane and intranasally challenged with 10 lethal dose 50 (LD_{50}) of mouse-adapted A/Netherlands/602/2009 H1N1 virus, with 10 µl of virus administered into each nostril (20 µl total). For therapeutic treatment of A/Netherlands/602/2009 and prophylactic treatment of A/Fort Monmouth/1/1947, mice were anaesthetized with a ketamine–xylazine–water cocktail (0.15 mg ketamine per kg and 0.03 mg per kg xylazine; 100 ml intraperitoneally) and infected with 10 LD_{50} of A/Netherlands/602/2009 or A/Fort Monmouth/1/1947. As a read out, survival and weight loss were monitored 1–2 times daily for 2 weeks. Mice were euthanized upon 25% weight loss or at the end of the experiment (14 days post-challenge). Five mice per condition per experiment with two biological replicates were utilized based on a previously performed power analysis. Data were pooled for analysis.

To determine differences in lung viral load, 5 mg per kg of the antibody cocktails was administered prophylactically as described above. Two hours after mAb administration, mice ($n = 5$ mice per group) were anaesthetized and intranasally challenged with 1 LD_{50} of A/Netherlands/602/2009. Lungs were collected at day 3 and day 6 post-infection, homogenized and viral load was determined via plaque assay. All experiments were done in accordance with the University of Chicago and Icahn School of Medicine at Mount Sinai Institutional Animal Care and Use Committees. Animals studies were not blinded or randomized.

Plaque assay

For determination of viral load in mouse lung tissues a standard plaque assay was performed. Confluent monolayers of MDCK cells were infected with serial dilutions of homogenized lung tissue ranging from 1:10 to 1:1,000,000 diluted in 1X MEM (1%

penicillin–streptomycin antibiotics mix, 1% HEPES, 1% l-glutamine and 1% sodium-bicarbonate (Gibco)) for 1 h at 33 °C, with shaking every 15 min. Afterwards, an overlay containing 2% Oxoid agar (Thermo Fisher), H₂O, 2X MEM, DEAE and TPCK-treated trypsin was added to the cells. The plates were incubated at 33 °C for 3 days and then fixed with 3.7% paraformaldehyde overnight at 4 °C. Plaques were visualized by immunostaining. Here, the agar overlay was removed and the plates blocked with 3% milk and PBS. The blocking solution was removed and primary antibody ((H1N1 guinea pig anti-sera (generated in house)) diluted 1:3,000 in 1% milk and PBS was added for 1 h. The plates were washed three times with PBS and secondary antibody (anti-mouse IgG H&L peroxidase-conjugated (Rockland) diluted 1:3,000 in 1% milk and PBS was added for 1 h. The plates were washed three times with PBS and developed by using KPL TrueBlue Peroxidase Substrate (SeraCare).

Antibody-dependent cellular cytotoxicity reporter assay

A549 cells were maintained in DMEM supplemented with 10% FBS, 10 U ml⁻¹ penicillin, and 10 mg ml⁻¹ streptomycin) and were plated in 96-well, white-walled plates (Costar) at 2.5×10^5 cells per ml overnight at 37 °C with 5% CO₂. The following day, cells were washed with PBS and infected with A/Netherlands/602/2009 at a multiplicity of infection of 5 in UltraMDCK media (Lonza) for 24 h in the absence of TPCK-treated trypsin. mAbs were serially diluted in assay buffer (RPMI 1640 supplemented with 4% ultra-low IgG FBS; Gibco), starting at 60 µg ml⁻¹ and diluted threefold. Cell medium was aspirated and 25 µl of assay buffer and 25 µl of diluted antibody were added to each well. Jurkat cells expressing human FcgRIIIa with a NFAT-driven luciferase reporter gene (Promega) were diluted to 3×10^6 cells per ml, 25 µl of cells was added to each well and incubated at 37 °C with 5% CO₂ for 6 h. Plates were removed from the incubator and placed at room temperature for 15 min. Of the BioGlo luciferase substrate (Promega), 75 µl was added to each well and luminescence was read immediately using a Syngery H1 hybrid multimode microplate reader (Biotek). EC₅₀ values were determined using Prism 8 (GraphPad).

HA–antibody binding footprint mapping

The footprints of three mAbs (FISW84 (PDB: 6HJQ), CR9114 (PDB: 4FQI) and FI6v3 (PDB: 3ZTN)) were mapped onto one HA protomer (A/California/4/2009, PDB: 4M4Y) using UCSF Chimera⁶⁰ and Adobe Photoshop. Negative-stain EM maps of HA–Fab complexes were aligned in UCSF Chimera and estimated footprints were mapped onto one HA protomer. Individual protomers of the HA trimer are indicated in different shades of grey.

Negative-stain EM

Immune complexes were prepared by incubating Fab with HA (A/California/04/2009 with E376K or E376G stabilizing mutations) at greater than 3:1 molar ratio for 2 h at room temperature. Samples were deposited at approximately 10 µg ml⁻¹ on glow-discharged, carbon-coated 400 mesh copper grids (Electron Microscopy Sciences) and stained with 2% w/v uranyl formate. Samples were imaged at ×52,000 magnification, 120 kV, on a Tecnai Spirit T12 microscope equipped with an Eagle CCD 4k camera (FEI) or ×62,000 magnification, 200 kV, on a Tecnai T20 microscope equipped with a CMOS 4k camera (TVIPS). Micrographs were collected with Leginon, single particles were processed with Appion, Relion and XQuartz, and footprints were mapped with UCSF Chimera, and figures were made with UCSF Chimera^{60,61,62,63}.

Cryo-EM

222-1C06 and 045-09 2B05 Fabs were incubated at greater than 3:1 molar ratio with HA (A/California/7/2009, E376K) for 1 h at room temperature. 045-09 2B05 Fab, targeting the lateral patch, was added to the immune complex to induce particle tumbling and increase angular sampling³. Using a Thermo Fisher Vitrobot, the immune complex (0.5 mg ml⁻¹) incubated with lauryl maltose neopentyl glycol (5 µM, Anatrace) was deposited onto glow-discharged Au 1.2/1.3 300 mesh grids (Electron Microscopy Sciences), blotted for 7 s, and plunge-frozen in liquid ethane. Samples were imaged at ×36,000 nominal magnification on a 200 kV Talos Arctica electron microscope (FEI) with a CETA 4k CMOS camera (FEI, total dose 49.92 e/Å²) and Gatan K2 Summit detector in counting mode. 2,243 micrographs were collected, aligned and CTF-corrected using Leginon, MotionCor2 in Appion, and Patch-CTF in CryoSPARC2, respectively^{61,62,64,65}. In CryoSPARC2, particles were picked using apo HA templates, selected through reference-free 2D classification, and reconstructed through 3D classification and refinement. The final map resolved to a global 3.75 Å resolution with C3 symmetry and 44,224 particles. Figures were made in Prism 8 (GraphPad) and UCSF Chimera⁶⁰.

Model building and refinement

A predicted model of 222-1C06 Fab was generated using abYsis (<http://www.abysis.org/abysis/>) and docked into EM density along with an initial model of CA09 H1 HA + 045-09 2B05 (PDB: 7MEM). The initial model was iteratively refined using COOT and Rosetta^{66,67}. The final model was numbered using the H3 and Kabat numbering schemes. The final model and map were evaluated using MolProbity, EMRinger^{68,69}, Phenix and the PDB validation server. After modelling the immune complex, we segmented the Fab density from HA in the cryo-EM map and mapped the footprint of the 222-1C06 model in the HA density. Cryo-EM data collection and refinement statistics are included in Extended Data Table 4.

Electron microscopy polyclonal epitope mapping

Human serum samples were heat-inactivated at 55 °C for 30 min before incubating on Capture Select IgG-Fc (ms) Affinity Matrix (Fisher) to bind IgG at 4 °C for 72 h on a rotator. Samples with IgG bound to resin were centrifuged at 4,000 rpm and supernatant was collected. IgG samples were washed three times with PBS followed by centrifugation to remove supernatant. Samples were buffer exchanged into buffer containing 100 mM Tris, 2 mM EDTA and 10 mM l-cysteine through centrifugation with Amicon filters, then incubated with papain for 4 h at 37 °C, shaking at 80 rpm. The digestion reactions were quenched with 50 mM iodoacetamide, buffer exchanged to TBS and separated by size-exclusion chromatography (SEC) with a Superdex 200 increase 10/300 column (GE Healthcare). Fab and undigested IgG were collected and concentrated, and 500 µg Fab was complexed with 10 µg HA for 18 h at room temperature. Reactions were purified by SEC and immune complexes were collected and concentrated. Negative-stain EM grids were prepared as described above.

Membrane-bound HA and mAb staining

HEK293T cells were plated into a six-well plate and transfected overnight with 0.2 µg of plasmid and 10 µg ml⁻¹ PEI. After 12–16 h, media were replaced with PFHM-II (Gibco) and cells were rested for 3 days. Transfected cells were trypsinized, washed and aliquoted. Cells were stained with 10 µg ml⁻¹ of individual mAbs for 30 min. Cells were washed and stained with anti-human IgG Fc-BV421 for 30 min. Cells were washed two times and run on a BD LSRLFortessa and collected with BD FACSDiva software. Data were analysed using FlowJo v10.

Single-cell RNA sequencing and repertoire analysis

ch5/1⁺ memory B cells (CD19⁺CD27⁺HA⁺) were bulk sorted and partitioned into nanolitre-scale gel bead-in-emulsions (GEMs) to achieve single-cell resolution using the 10x Genomics Chromium Controller and according to the manufacturer's instruction (10x Genomics). The sorted single cells were processed according to 5' gene expression and B cell immunoglobulin enrichment instruction to prepare the libraries for sequencing. Libraries were sequenced using an Illumina HiSeq 4000 at Northwestern University or an Illumina NextSeq 500 at the University of Chicago. Cellranger Single-Cell Software Suite (version 3.0) was used to perform sample demultiplexing, barcode processing, and single-cell 5' and V(D)J counting, and Cellranger mkfastq was used to demultiplex raw base call (BCL) files into sample-specific fastq files. Subsequently, GRCh38-1.2.0 and cellranger-vdj-GRCh38-alta-ensembl-2.0.0 were used as references for the transcriptome and V(D)J assembly, respectively. Cellranger counts and Cellranger vdj package were used to identify gene expression and assemble V(D)J pairs of antibodies.

Single-cell datasets were analysed using Seurat 3 toolkit (Version 3.2.0). We performed conventional pre-process steps for all 20 donors including cell quality control (QC), normalization, identification of highly variable features, data scaling and linear dimensional reduction. More specifically, we only kept cells with more than 200 and less than 2,500 detected genes for the QC step. We also filtered out cells with high mitochondrial gene expression using a ‘softThreshold’ function in the R package LinQ-View (version 0.99)⁷⁰. We normalized the RNA data using conventional log normalization. We identified 2,000 highly variable genes for each dataset and performed principle component analysis (PCA) in linear dimensional reduction step. We then integrated all 20 single-cell datasets from vaccinated participants to remove batch effects using the Anchor method in Seurat 3. In this analysis, we filtered our dataset and only kept cells with both transcriptome and full length and paired heavy and light chain V(D)J sequences ($n = 1,952$). From these cells, we identified a group of ‘VH1-69/κ’ B cells that used the VH1-69 gene and κ-light chain, which is enriched for B cells targeting the BN stalk epitope. We also identified a group of ‘anchor epitope’-specific B cells by the following rules: (1) VH locus: VH3-23, VH3-30, VH3-30-3 or VH3-48; (2) VK locus: VK3-11 or VK3-15; (3) JK locus: JK4 or JK5; (4) K-CDR3 length equal to 10; and (5) a ‘NWP’ pattern in the K-CDR3 peptide.

HA conservation modelling

Pan-H1 conservation models are based on consensus strains (listed in Supplementary Table 4) of distinct H1 clades isolated from humans, swine and avian sources, as described in Zhuang et al.⁷¹ and inclusion of the Eurasian swine-like A/swine/Jiangsu/J004/2018 (ref. ²³). To generate the group 1 HA conservation model, we selected one representative sequence for each group 1 HA subtype from FluDB (<https://www.fludb.org/>; Supplementary Table 5) according to a previous study⁷². A multiple sequence alignment from these HA protein sequences was generated using MUSCLE⁷³ and the conservation of each residue was quantified using an entropy model⁴¹. Seasonal H1 conservation models are based on consensus strains of H1N1 viruses (59 strains total) circulating between 1918–1957 and 1976–2019, which was previously described³. Amino acid alignments and H3 numbering were performed using Librator⁷⁴ and Burke and Smith HA numbering⁷².

Structure prediction

To predict the structures of the investigated Fv fragments (222-1C06, FISW84, 241 IgA 2F04 and SFV009 3G01) with A/California/4/2009 E47G HA (PDB: 7MEM), we applied the program RosettaAntibody^{67,75,76}. The Fvs were protonated using the Protonate 3D tool^{77,78}. Charge neutrality was ensured by utilizing the uniform background plasma approach in AMBER^{79,80}. Using the tleap tool of the

AmberTools20 (ref. ⁸¹) package, the structure models were soaked in cubic water boxes of TIP3P water molecules with a minimum wall distance of 10 Å to the protein⁸². Parameters for all antibody models derive from the AMBER force field 14SB⁸³. The Fvs were carefully equilibrated using a multistep equilibration protocol⁸⁴.

Metadynamics simulations

To enhance the sampling of the conformational space, well-tempered bias-exchange metadynamics^{85,86,87} simulations were performed in GROMACS^{88,89} with the PLUMED 2 implementation⁹⁰. We chose metadynamics as it enhances sampling on predefined collective variables. The sampling is accelerated by a history-dependent bias potential, which is constructed in the space of the collective variables^{85,86,91}. As collective variables, we used a well-established protocol, boosting a linear combination of sine and cosine of the ψ torsion angles of all six CDR loops calculated with functions MATHEVAL and COMBINE implemented in PLUMED 2 (ref. ⁹⁰). As discussed previously, the ψ torsion angle captures conformational transitions comprehensively⁹². The underlying method presented in this paper has been validated in various studies against a large number of experimental results⁹³. The simulations were performed at 300 K in an NpT ensemble using the GPU implementation of the pmemd module⁹⁴ to be as close to the experimental conditions as possible and to obtain the correct density distributions of both protein and water. We used a Gaussian height of 10.0 kJ mol⁻¹ and a width of 0.3 rad. Gaussian deposition occurred every 1,000 steps and a biasfactor of 10 was used. 500 ns of bias-exchange metadynamics simulations were performed for the prepared Fv structures. The resulting trajectories were aligned to the whole Fv and clustered with the program cpptraj^{80,95} using the average linkage hierarchical clustering algorithm with a RMSD cut-off criterion of 1.2 Å resulting in a large number of clusters. The cluster representatives for the antibody fragments were equilibrated and simulated for 100 ns using the AMBER 20 (ref. ⁸¹) simulation package.

Molecular dynamics simulations

Molecular dynamics simulations were performed in an NpT ensemble using the pmemd.cuda module of AMBER 20 (ref. ⁸⁰). Bonds involving hydrogen atoms were restrained with the SHAKE algorithm⁹⁶, allowing a time step of 2.0 fs. Atmospheric pressure (1 bar) of the system was set by weak coupling to an external bath using the Berendsen algorithm⁹⁷. The Langevin thermostat⁹⁸ was used to maintain the temperature during simulations at 300 K.

With the obtained trajectories, we performed a time-lagged independent component analysis (tICA) using the Python library PyEMMA 2, using a lag time of 10 ns. tICA

was applied to identify the slowest movements of the investigated Fab fragments and consequently to obtain a kinetic discretization of the sampled conformational space⁹⁹. On the basis of the tICA conformational spaces, thermodynamics and kinetics were calculated with a Markov-state model¹⁰⁰ of all six CDR loops by using PyEMMA 2. The resulting kinetically dominant ensemble in solution was further used to predict the interactions of H1 with the Fvs. To model the complex and to predict interactions in the binding interface, we used the crystal structure of the full-length influenza HA (PDB: 7MEM) as template structure. In addition, the obtained complex structure was further minimized and equilibrated.

Statistical analysis

All statistical analyses were performed using Prism software (GraphPad versions 8 and 9) or R. Sample sizes (n) for the number of mAbs tested are indicated in corresponding figures or in the centre of pie graphs. The number of biological repeats for experiments and specific tests for statistical significance used are indicated in the corresponding figure legends. P values less than or equal to 0.05 were considered significant: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ and **** $P < 0.0001$.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Repertoire data generated from single-cell RNA sequencing data were deposited at Mendeley Data (<https://data.mendeley.com/datasets/jzsx489pmk/1>). Accession numbers for all other anchor-targeting mAbs are included in Supplementary Table 6. Electron microscopy maps were deposited to the Electron Microscopy Data Bank under accession IDs: [EMD-25634](#)–[EMD-25646](#). The cryo-EM map and model of anchor and lateral patch Fabs binding H1 HA were deposited to the RCSB database with accession numbers [EMD-25655](#)/PDB [7T3D](#). All next-generation sequencing data for 045-09 2B06 deep mutational scanning and for the H1N1 mutational scanning can be found on the Sequence Read Archive under BioProject accession number [PRJNA309339](#). The following Protein Data Bank accession numbers were downloaded and included in the paper: [6HJQ](#), [3SDY](#), [4NM8](#), [4M4Y](#), [4WE4](#), [4JTV](#), [4FQI](#), [3ZTN](#) and [7MEM](#). Sera from the vaccine cohorts are unique to this study and are not publicly available. All source data are included with the paper. All other material is available on reasonable request to the corresponding authors. [Source data](#) are provided with this paper.

References

1. Paules, C. I., Marston, H. D., Eisinger, R. W., Baltimore, D. & Fauci, A. S. The pathway to a universal influenza vaccine. *Immunity* **47**, 599–603 (2017).
2. Park, J. K. et al. Pre-existing immunity to influenza virus hemagglutinin stalk might drive selection for antibody-escape mutant viruses in a human challenge model. *Nat. Med.* **26**, 1240–1246 (2020).
3. Guthmiller, J. J. et al. First exposure to the pandemic H1N1 virus induced broadly neutralizing antibodies targeting hemagglutinin head epitopes. *Sci. Transl. Med.* **13**, eabg453 (2021).

4. Bernstein, D. I. et al. Immunogenicity of chimeric haemagglutinin-based, universal influenza virus vaccine candidates: interim results of a randomised, placebo-controlled, phase 1 clinical trial. *Lancet Infect. Dis.* **20**, 80–91 (2020).
5. Nachbagauer, R. et al. A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. *Nat. Med.* **27**, 106–114 (2020).
6. Ng, S. et al. Novel correlates of protection against pandemic H1N1 influenza A virus infection. *Nat. Med.* **25**, 962–967 (2019).
7. Aydillo, T. et al. Pre-existing hemagglutinin stalk antibodies correlate with protection of lower respiratory symptoms in flu-infected transplant patients. *Cell Rep. Med.* **1**, 100130 (2020).
8. Wrammert, J. et al. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J. Exp. Med.* **208**, 181–193 (2011).
9. Andrews, S. F. et al. Immune history profoundly affects broadly protective B cell responses to influenza. *Sci. Transl. Med.* **7**, 316ra192 (2015).
10. Dreyfus, C. et al. Highly conserved protective epitopes on influenza B viruses. *Science* **337**, 1343–1348 (2012).
11. Benton, D. J. et al. Influenza hemagglutinin membrane anchor. *Proc. Natl Acad. Sci. USA* **115**, 10112–10117 (2018).
12. Ekiert, D. C. et al. A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* **333**, 843–850 (2011).
13. Friesen, R. H. et al. A common solution to group 2 influenza virus neutralization. *Proc. Natl Acad. Sci. USA* **111**, 445–450 (2014).

14. Corti, D. et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850–856 (2011).
15. DiLillo, D. J., Palese, P., Wilson, P. C. & Ravetch, J. V. Broadly neutralizing anti-influenza antibodies require Fc receptor engagement for in vivo protection. *J. Clin. Invest.* **126**, 605–610 (2016).
16. He, W. et al. Epitope specificity plays a critical role in regulating antibody-dependent cell-mediated cytotoxicity against influenza A virus. *Proc. Natl Acad. Sci. USA* **113**, 11931–11936 (2016).
17. Guthmiller, J. J. et al. Polyreactive broadly neutralizing B cells are selected to provide defense against pandemic threat influenza viruses. *Immunity* **53**, 1230–1244.e5 (2020).
18. Sutton, T. C. et al. In vitro neutralization is not predictive of prophylactic efficacy of broadly neutralizing monoclonal antibodies CR6261 and CR9114 against lethal H2 influenza virus challenge in mice. *J. Virol.* **91**, e01603-17 (2017).
19. Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150–153 (2020).
20. Han, J. et al. Polyclonal epitope mapping reveals temporal dynamics and diversity of human antibody responses to H5N1 vaccination. *Cell Rep.* **34**, 108682 (2021).
21. Impagliazzo, A. et al. A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science* **349**, 1301–1306 (2015).
22. Yassine, H. M. et al. Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat. Med.* **21**, 1065–1070 (2015).
23. Sun, H. et al. Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. *Proc. Natl Acad. Sci. USA* **117**, 17204–17210(2020).

24. Caskey, M. et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522**, 487–491 (2015).
25. Caskey, M. et al. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nat. Med.* **23**, 185–191 (2017).
26. Bar, K. J. et al. Effect of HIV antibody VRC01 on viral rebound after treatment interruption. *N. Engl. J. Med.* **375**, 2037–2050 (2016).
27. Mendoza, P. et al. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* **561**, 479–484 (2018).
28. Raymond, D. D. et al. Conserved epitope on influenza-virus hemagglutinin head defined by a vaccine-induced antibody. *Proc. Natl Acad. Sci. USA* **115**, 168–173 (2018).
29. Linderman, S. L. et al. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013–2014 influenza season. *Proc. Natl Acad. Sci. USA* **111**, 15798–15803 (2014).
30. Tate, M. D. et al. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* **6**, 1294–1316 (2014).
31. Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl Acad. Sci. USA* **114**, E7348–E7357 (2017).
32. Dugan, H. L. et al. Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans. *Sci. Transl. Med.* **12**, eabd3601 (2020).
33. van der Lubbe, J. E. M. et al. Mini-hemagglutinin vaccination induces cross-reactive antibodies in pre-exposed NHP that protect mice against lethal influenza challenge. *NPJ Vaccines* **3**, 25 (2018).
34. Cantisani, R. et al. Vaccine adjuvant MF59 promotes retention of unprocessed antigen in lymph node macrophage compartments and

follicular dendritic cells. *J. Immunol.* **194**, 1717–1725 (2015).

35. Liang, F. et al. Vaccine priming is restricted to draining lymph nodes and controlled by adjuvant-mediated antigen uptake. *Sci. Transl. Med.* **9**, eaal2094 (2017).
36. Khurana, S. et al. Vaccines with MF59 adjuvant expand the antibody repertoire to target protective sites of pandemic avian H5N1 influenza virus. *Sci. Transl. Med.* **2**, 15ra15 (2010).
37. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**, 1272–1275 (2010).
38. Guthmiller, J. J., Dugan, H. L., Neu, K. E., Lan, L. Y. & Wilson, P. C. An efficient method to generate monoclonal antibodies from human B cells. *Methods Mol. Biol.* **1904**, 109–145 (2019).
39. Wrammert, J. et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667–671 (2008).
40. Smith, K. et al. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat. Protoc.* **4**, 372–384 (2009).
41. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
42. Doud, M. B., Lee, J. M. & Bloom, J. D. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nat. Commun.* **9**, 1386 (2018).
43. Lang, S. et al. Antibody 27F3 broadly targets influenza A group 1 and 2 hemagglutinins through a further variation in VH1-69 antibody orientation on the HA stem. *Cell Rep.* **20**, 2935–2943 (2017).
44. Xu, R. et al. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009

- H1N1 influenza pandemic. *J. Virol.* **86**, 9221–9232 (2012).
45. Fu, Y. et al. A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve. *Nat. Commun.* **7**, 12780 (2016).
 46. Wu, N. C. et al. Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses. *Science* **368**, 1335–1340 (2020).
 47. Tan, G. S. et al. A pan-H1 anti-hemagglutinin monoclonal antibody with potent broad-spectrum efficacy in vivo. *J. Virol.* **86**, 6179–6188 (2012).
 48. Yassine, H. M. et al. Use of hemagglutinin stem probes demonstrate prevalence of broadly reactive group 1 influenza antibodies in human sera. *Sci. Rep.* **8**, 8628 (2018).
 49. Lingwood, D. et al. Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* **489**, 566–570 (2012).
 50. Cotter, C. R., Jin, H. & Chen, Z. A single amino acid in the stalk region of the H1N1pdm influenza virus HA protein affects viral fusion, stability and infectivity. *PLoS Pathog.* **10**, e1003831 (2014).
 51. Sui, J. et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
 52. Vanderlinden, E. et al. Novel inhibitors of influenza virus fusion: structure–activity relationship and interaction with the viral hemagglutinin. *J. Virol.* **84**, 4277–4288 (2010).
 53. Hai, R. et al. Influenza viruses expressing chimeric hemagglutinins: globular head and stalk domains derived from different subtypes. *J. Virol.* **86**, 5774–5781 (2012).

54. Shlomchik, M. J., Aucoin, A. H., Pisetsky, D. S. & Weigert, M. G. Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. *Proc. Natl Acad. Sci. USA* **84**, 9150–9154 (1987).
55. Doud, M. B. & Bloom, J. D. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* **8**, 155 (2016).
56. Doud, M. B., Hensley, S. E. & Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **13**, e1006271 (2017).
57. Bloom, J. D. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* **16**, 168 (2015).
58. Henry Dunand, C. J. et al. Preexisting human antibodies neutralize recently emerged H7N9 influenza strains. *J. Clin. Invest.* **125**, 1255–1268 (2015).
59. Chen, Y. Q. et al. Influenza infection in humans induces broadly cross-reactive and protective neuraminidase-reactive antibodies. *Cell* **173**, 417–429.e10 (2018).
60. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
61. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
62. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
63. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

64. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
65. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
66. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
67. Weitzner, B. D. et al. Modeling and docking of antibody structures with Rosetta. *Nat. Protoc.* **12**, 401–416 (2017).
68. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
69. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
70. Li, L. et al. Improved integration of single-cell transcriptome and surface protein expression by LinQ-View. *Cell Rep. Methods* **1**, 100056 (2021).
71. Zhuang, Q. et al. Diversity and distribution of type A influenza viruses: an updated panorama analysis based on protein sequences. *Virol. J.* **16**, 85 (2019).
72. Burke, D. F. & Smith, D. J. A recommended numbering scheme for influenza A HA subtypes. *PLoS ONE* **9**, e112302 (2014).
73. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

74. Li, L. et al. Librator, a platform for optimized sequence editing, design, and expression of influenza virus proteins. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.29.441999> (2021).
75. Weitzner, B. D., Kuroda, D., Marze, N., Xu, J. & Gray, J. J. Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization. *Proteins* **82**, 1611–1623 (2014).
76. Stein, A. & Kortemme, T. Improvements to robotics-inspired conformational sampling in Rosetta. *PLoS ONE* **8**, e63090 (2013).
77. Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **75**, 187–205 (2009).
78. Molecular Operating Environment (MOE). (2020).
79. Hub, J. S., de Groot, B. L., Grubmuller, H. & Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **10**, 381–390 (2014).
80. Roe, D. R. & Cheatham, T. E. 3rd PTraj and CPPtraj: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
81. Case, D. A. et al. AMBER 2020 (University of California, San Francisco, 2020).
82. 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).
83. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

84. Wallnoefer, H. G., Liedl, K. R. & Fox, T. A challenging system: free energy prediction for factor Xa. *J. Comput. Chem.* **32**, 1743–1752 (2011).
85. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *WIREs Computat. Mol. Sci.* **1**, 826–843 (2011).
86. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
87. Biswas, M., Lickert, B. & Stock, G. Metadynamics enhanced Markov modeling of protein dynamics. *J. Phys. Chem. B* **122**, 5508–5514 (2018).
88. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
89. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
90. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G. PLUMED 2: new feathers for an old bird. *Comp. Phys. Commun.* **185**, 604–613 (2014).
91. Ilott, A. J., Palucha, S., Hodgkinson, P. & Wilson, M. R. Well-tempered metadynamics as a tool for characterizing multi-component, crystalline molecular machines. *J. Phys. Chem. B* **117**, 12286–12295 (2013).
92. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
93. Fernandez-Quintero, M. L. et al. Antibodies exhibit multiple paratope states influencing VH-VL domain orientations. *Commun. Biol.* **3**, 589

(2020).

94. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
95. Shao, J., Tanner, S. W., Thompson, N. & Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **3**, 2312–2334 (2007).
96. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
97. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
98. Adelman, S. A. & Doll, J. D. Generalized Langevin equation approach for atom/solid-surface scattering: general formulation for classical scattering off harmonic solids. *J. Chem. Phys.* **64**, 2375–2388 (1976).
99. Scherer, M. K. et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
100. Chodera, J. D. & Noe, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).

Acknowledgements

We thank all of the volunteers who participated in this study; S. Andrews, R. Ahmed, J. Wrammert and K. Neu for initiating studies on the 2009 MIV, 2010 TIV and 2014 QIV cohorts; C. Mariottini, J. Feser, D. Stadlbauer and

A.-K. Palm for their help on the cHA vaccine trial; I. Wilson for fruitful discussion and feedback on experimental design; B. Anderson and H. Turner for assistance with electron microscopy; and the teams at PATH, GSK, Cincinnati Children's Hospital Medical Center, and Duke University for their work on the cHA vaccine trial (NCT03300050). This project was funded in part by the National Institute of Allergy and Infectious Diseases (NIAID; National Institutes of Health grant numbers K99AI159136 (to J.J.G.), U19AI082724 (to P.C.W.), U19AI109946 (to P.C.W.), U19AI057266 (to P.C.W.), P01AI097092 (to P.P.), R01AI145870-01 (to P.P.), R21AI146529 (to L.C.) and T32AI007244-36 (to J.H.), the NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS) grant number HHSN272201400005C (to P.C.W.), HHSN272201400008C (to L.C., F.K., A.G.-S. and P.P.), and the NIAID Centers of Excellence for Influenza Research and Response (CEIRR) grant number 75N93019R00028 (to P.C.W., F.K., A.G.-S. and P.P.). This work was also partially supported by the NIAID Collaborative Influenza Vaccine Innovation Centers (CIVIC; 75N93019C00051, to F.K., A.G.-S., P.P., A.B.W. and P.C.W.). The cHA vaccine trial and evaluation of immunity thereafter was funded in part by the Bill and Melinda Gates Foundation (OPP1084518). The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Gates Foundation. This study was also funded by the Austrian Science Fund grant number P34518 (to M.L.F.-Q. and K.R.L.).

Author information

Author notes

1. Carole Henry, Alec W. Freyn & Raffael Nachbagauer

Present address: Moderna Inc., Cambridge, MA, USA

2. Patrick C. Wilson

Present address: Drukier Institute for Children's Health, Department of Pediatrics, Weill Cornell Medicine, New York, NY, USA

3. These authors contributed equally: Jenna J. Guthmiller, Julianna Han

Affiliations

1. Department of Medicine, Section of Rheumatology, University of Chicago, Chicago, IL, USA

Jenna J. Guthmiller, Henry A. Utset, Lei Li, Carole Henry, Olivia Stovicek, Nai-Ying Zheng, Micah E. Tepora, Dalia J. Bitar, Siriruk Changrob, Min Huang & Patrick C. Wilson

2. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA

Julianna Han, Sara T. Richey, Alba Torrents de la Peña & Andrew B. Ward

3. Committee on Immunology, University of Chicago, Chicago, IL, USA

Linda Yu-Ling Lan, Christopher T. Stamper, Haley L. Dugan & Patrick C. Wilson

4. Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Meagan McMahon, George O'Dell, Alec W. Freyn, Fatima Amanat, Victoria Rosado, Shirin Strohmeier, Adolfo García-Sastre, Raffael Nachbagauer, Peter Palese & Florian Krammer

5. Center for Molecular Biosciences Innsbruck, Department of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria

Monica L. Fernández-Quintero & Klaus R. Liedl

6. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Lauren Gentles & Jesse D. Bloom

7. Department of Microbiology, University of Washington, Seattle, WA, USA

Lauren Gentles & Jesse D. Bloom

8. Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Adolfo García-Sastre & Peter Palese

9. Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Adolfo García-Sastre

10. The Tisch Cancer Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Adolfo García-Sastre

11. Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Adolfo García-Sastre

12. Department of Genome Sciences, University of Washington, Seattle, WA, USA

Jesse D. Bloom

13. Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Jesse D. Bloom

14. Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA

Lynda Coughlan

15. Center for Vaccine Development and Global Health (CVD), University of Maryland School of Medicine, Baltimore, MD, USA

Lynda Coughlan

Contributions

J.J.G., J.H., A.B.W. and P.C.W. conceptualized the study. J.J.G., J.H., H.A.U., L.L., L.Y.-L.L., A.W.F., M.L.F.-Q., C.H., C.T.S., M.M., G.O., F.A., L.G., N.-Y.Z., S.S., M.H., J.D.B. and L.C. developed the methodology. J.J.G., J.H., H.A.U., L.L., L.Y.-L.L., A.W.F., M.L.F.-Q., C.H., C.T.S., M.M., G.O., V.R., F.A., O.S., L.G., H.L.D., S.T.R., A.T.d.l.P., M.E.T., D.J.B., S.C. and L.C. conducted the investigation. J.J.G., J.H., H.A.U., L.L., M.L.F.-Q., M.M., L.G. and S.T.R. performed visualization. J.J.G., J.H., A.G.-S., P.P., F.K., L.C., A.B.W. and P.C.W. acquired funding. Project administration was done by J.J.G., J.H., C.H., A.G.-S., K.R.L., R.N., P.P., F.K., A.B.W. and P.C.W. J.J.G., J.H., A.B.W. and P.C.W. supervised the study. J.J.G. and J.H. wrote the original draft of the manuscript. J.J.G., J.H., H.A.U., L.L., L.Y.-L.L., C.H., C.T.S., M.M., G.O., A.W.F., O.S., L.G., H.L.D., N.-Y.Z., S.T.R., A.T.d.l.P., M.E.T., D.J.B., S.C., S.S., M.H., A.G.-S., R.N., P.P., J.D.B., F.K., L.C., A.B.W. and P.C.W. reviewed and edited the manuscript.

Corresponding authors

Correspondence to [Jenna J. Guthmiller](#), [Andrew B. Ward](#) or [Patrick C. Wilson](#).

Ethics declarations

Competing interests

The Icahn School of Medicine at Mount Sinai has patents (10736956, 10583188, 10137189, 10131695, 9968670, 9371366) and has submitted patent applications (10736956, 10583188, 10137189, 10131695, 9968670,

9371366) on universal influenza virus vaccines naming A.G.-S., R.N., P.P. and F.K. as inventors. The University of Chicago has submitted patent applications (632292804) on anti-anchor mAbs naming J.J.G. and P.C.W. as inventors. The A.G.-S. laboratory has received research support from Pfizer, Senhwa Biosciences, Kenall Manufacturing, Avimex, Johnson & Johnson, Dynavax, 7Hills Pharma, Pharmamar, ImmunityBio, Accurius, Nanocomposix, Hexamer, N-fold LLC and Merck, outside of the reported work. A.G.-S. has consulting agreements for the following companies involving cash and/or stock: Vivaldi Biosciences, Contrafect, 7Hills Pharma, Avimex, Vaxalto, Pagoda, Accurius, Esperovax, Farmak, Applied Biological Laboratories and Pfizer, outside of the reported work.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

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 Binding and neutralization features of anchor epitope-binding mAbs. Related to Fig. 1.](#)

a, Proportion of HA⁺ mAbs binding to distinct HA domains (left) and proportion of stalk-binding mAbs binding the CS domain (right). Number in the center of the pie graphs represent the number of mAbs tested. **b**, Proportion of mAbs per cohort that bind the HA stalk domain. **c**, Negative stain 2D class averages of 047-09 4F04, 241 IgA 2F04, and 222-1C06 binding to H1 (A/California/7/2009 E376K HA). Imaging of 047-09 4F04 was performed at ×52,000 normal magnification and of 222-1C06 and 241 IgA 2F04 at ×62,000 normal magnification. **d**, Overlay of 047-09 4F04, 241 IgA 2F04, 222-1C06, and FISW84 (PDB:6HJQ) Fabs binding the anchor

epitope of A/California/4/2009 HA. **e**, Overlay of CR8020 (PDB:3SDY), CR8043 (PDB:4NM8), and FISW84 (PDB:6HJQ) modeled on A/California/7/2009 E376G (PDB:4M4Y). **f**, Footprints of anchor mAb 222-1C06 on H1 (top; PDB: 4M4Y) and CR8020 and CR8043 on H3 (bottom; PDB:4WE4). **g**, Heatmap of apparent affinity (K_d ; M) of anchor-targeting mAbs binding to historical and recent H1N1 viruses. **h**, Neutralization potency of anchor-binding mAbs ($n = 15$) against H1-expressing viruses. **i**, Representative microneutralization curves of anchor- ($n = 42$) and CS-binding ($n = 29$) mAbs against A/California/7/2009. **j**, IC_{80} of anchor- and CS-binding mAbs against A/California/7/2009. **k**, ADCC activity of mAbs targeting the CS and anchor epitopes. Dashed line represents the limit of detection (L.O.D.). **l**, ADCC potency of mAbs targeting the anchor ($n = 18$ mAbs) and CS ($n = 8$ mAbs) epitopes. Data in **h**, **j**, and **l** are represented as mean \pm S.D. Data in **j** and **l** were analyzed using a two-tailed unpaired non-parametric Mann-Whitney test.

[Source data](#)

[Extended Data Fig. 2 Anchor-targeting mAb binding to influenza subtype, viral mutants, and polyreactivity antigen panel. Related to Fig. 1.](#)

a, Proportion of anchor- ($n = 50$ mAbs) and CS-targeting mAbs ($n = 37$) binding influenza B viruses and H3N2 viruses. **b**, Negative stain 2D class averages ($\times 62,000$ normal magnification) of 222-1C06 binding to H2 (A/Singapore/1/1957), and H5 (A/Indonesia/5/2005). **c**, H2N2 neutralizing data of anchor- ($n = 11$ mAbs) and CS-binding mAbs ($n = 4$) represented as minimum neutralizing concentration. The limit of detection (L.O.D.) is 30 mg/ml. **d**, Proportion of mAbs targeting the anchor ($n = 50$ mAbs) or CS ($n = 50$ mAbs) epitope that are polyreactive. **e**, LPS binding strength, represented as area under the curve (AUC), of polyreactive mAbs targeting the anchor ($n = 30$ mAbs) and central stalk ($n = 43$ mAbs) epitopes. Data are mean \pm S.D. **f**, **g**, Anchor- and CS epitope-binding mAbs were tested for binding to A/California/7/2009 HA with naturally occurring and experimentally determined mutations induced by 045-09 2B06, a CS-binding mAb. **f**, Location of mutations modeled on A/California/4/2009 HA

(PDB: 4JTV). Residues in blue are located on HA1 and residues in red are located on HA2. Outlines represent binding footprints of 047-09 4F04 (sky blue) and CR9114 (green). **g**, Heatmap of mAb binding to WT and mutant HAs shown as the proportion of signal relative to mAb binding to the WT HA. Data in **a** and **d** were analyzed using Fisher's Exact tests. Data in **e** were analyzed using a two-tailed unpaired non-parametric Mann-Whitney test.

[Source data](#)

Extended Data Fig. 3 Anchor epitope-targeting mAbs are potently protective *in vivo* and lack ADCC activity. Related to Fig. 1.

a, b, Mice were prophylactically (2 h prior to infection; **a**) or therapeutically (48 h after infection; **b**) administered i.p. a cocktail of mAbs ($n = 5$ mAbs/cocktail) against the anchor- or CS-, or an anthrax-specific antibody. Mice were infected with 10 LD_{50} of A/Netherlands/602/2009 H1N1. Weight loss (top) and survival (bottom) of mice in each treatment group. **c**, Lung viral titers of mice in each prophylactic treatment group infected with 1 LD_{50} of A/Netherlands/602/2009. dpi, days post infection. **d**, Mice were prophylactically (2 h prior to infection) administered i.p. a cocktail of mAbs ($n = 5$ mAbs/cocktail) against the anchor- or CS-, or an anthrax-specific antibody. Mice were infected with 10 LD_{50} of A/Fort Monmouth/1/1947 H1N1. Weight loss (top) and survival (bottom) of mice in each treatment group. For **a**, **b**, and **d**, 10 mice per treatment group were used and data are pooled from two independent experiments. For **c**, 5 mice per treatment group and timepoint were used except for anchor cocktail group at dpi 3 only 4 mice were used. Data in **a**, **b**, and **d** are represented as mean \pm S.D and data in **c** are represented by mean \pm S.E.M. Kaplan-Meier curves in **a**, **b**, **d** were analyzed using a Log-rank Mantel-Cox test, and data in **c** were analyzed using multiple two-tailed unpaired non-parametric Kruskal-Wallis tests.

[Source data](#)

Extended Data Fig. 4 Additional repertoire and structural features of mAbs binding the anchor epitope. Related to Fig. 2.

a, VH locus usage by anchor- ($n = 52$ mAbs) and CS-binding mAbs ($n = 37$ mAbs). **b**, VH1 gene usage of mAbs targeting the CS epitope. **c**, VK locus usage by anchor- ($n = 52$ mAbs) and CS-binding mAbs ($n = 37$ mAbs). **d**, JK gene usage by anchor epitope-binding mAbs. **e**, Clonal expansions of anchor epitope-targeting mAbs. Numbers indicate heavy and light chain parings, which are described in Extended Data Table 2. **f**, Heavy and light chain sequences of the public clone. **g**, **h**, Mutations (**g**) and CDR3 amino acid (AA) lengths (**h**) of heavy and light chains of mAbs binding the anchor ($n = 52$ mabs) or CS ($n = 37$ mAbs) epitopes. Data are mean \pm S.D. **i**, Cryo-EM map of 222-1C06 binding to A/California/7/2009 E376K HA. **j**, **k**, Local resolution (**j**) and Fourier Shell Correlation (**k**) of 222-1C06 binding to HA. **l**, Aromatic pockets of 222-1C06 binding A/California/7/2009 E376K and FISW84 binding to A/duck/Alberta/35/1976 (PDB:6HJQ; top) and overlay of epitope:paratope interaction (bottom). **m**, MD simulations demonstrating the K-CDR3 NWP and H-CDR2 Y58 motifs of 222-1C06, FISW84, 241 IgA 2F04, and SFV009 3G01 binding to HA A/California/7/2009 HA. For left-hand panels in **l** and all panels in **m**, HA epitope contact residues (maroon) and heavy chain (green) and light chain (yellow) antibody contact residues of anchor mAb paratopes. Peach highlighted amino acids represent the fusion peptide of HA2. **n**, Fab-Fab interactions of the aromatic pocket of 222-1C06. **o**, MD simulation of the paratope flexibility of 222-1C06, highlighting the p-stacking of H-CDR2 and K-CDR3. **p**, Conservation of side-chain contacts of 222-1C06 across seasonal human H1N1 viruses circulating between 1918-2019. **q**, Deep mutational scanning of the side-chain contacts of 222-1C06. Data in **a** and **c** were analyzed using a Chi-square test, and data in **g**, **h** were analyzed by two-tailed unpaired non-parametric Mann-Whitney tests.

[Source data](#)

Extended Data Fig. 5 Features of anchor-targeting MBCs and EMPEM 2D classes. Related to Fig. 3.

a, 33 mAbs with anchor epitope-binding mAb repertoire features were generated and tested for competing for binding with 047-09 4F04. **b, c**, Number of heavy chain mutations (**b**) and isotype usage (**c**) of B cells with repertoire features of anchor-binding mAbs ($n = 119$ cells) or utilize VH1-69/kappa ($n = 365$ cells). **d, e**, 2D class averages of pAbs from donors 236 (**d**) and 241 (**e**) at days 7 and 14 post immunization binding to A/Michigan/45/2015 HA ($\times 62,000$ normal magnification). The last row of 2D classes in **d** is HA monomer complexes processed independently from trimer complexes. Data in **b** are represented as mean \pm S.D. Data in **b** were analyzed using a two-tailed paired non-parametric Wilcoxon matched-pairs signed rank test.

[Source data](#)

[Extended Data Fig. 6 Serum antibody kinetics of anchor- and CS-epitope binding antibodies after cHA vaccination and mAb binding to recombinant HAs. Related to Fig. 4.](#)

a, b, EC₅₀s of serum antibodies competing for binding with 047-09 4F04 for binding to the anchor epitope (**a**) and CR9114 for binding to the CS epitope (**b**). **a, b**, Kinetics of serum antibody responses against the anchor (**a**) and CS (**b**) epitopes. Data are mean \pm S.D. **c, d**, Proportion of stalk⁺ mAbs per donor (**c**) or proportion of donors with an isolated anchor mAb (**d**) upon first exposure to the pH1N1 virus (2009 MIV cohort) relative to donors who have repeatedly been exposed to pH1N1 (2010 TIV and 2014 QIV). Data in **c** are mean \pm S.D. Data in **c** includes only donors with an isolated anti-stalk mAb, whereas **d** includes all donors. **e, f**, Antibody titers (EC₅₀) of serum antibodies collected on day 113 and day 420 against the anchor epitope (**e**) and the CS epitope (**f**). Lines connect titers from the same donor and each pair of symbols represents one donor. **g**, Proportion of anchor epitope-binding mAbs binding to cHA (cH6/1) or mini-HA ($n = 50$). **h**, Representative flow cytometry plots of mAbs binding to A/California/7/2009 Cal09 HA and mini-HA (left) and geometric mean fluorescence intensity (gMFI) of mAbs binding to Cal09 and mini-HA (right). Data represent the mean \pm S.D. and each symbol represents an individual mAb. **i**, Proportion of anchor epitope-targeting mAbs binding to

A/California/7/2009 recombinant HA with a GCN4 or fibritin trimerization domain ($n = 50$). For data in **a**, **b**, **e**, and **f**, Group 1 $n = 10$ participants, group 2 $n = 7$ participants, group 4 $n = 7$ participants, group 3&5 $n = 6$. For data in **c**, first exposure $n = 7$ donors and repeated exposure $n = 4$ donors. For data in **d**, first exposure $n = 10$ donors and repeated exposure $n = 13$ donors. Data in **a**, **b** were analyzed using a two-tailed two-way ANOVA testing for simple effects within rows, data in **c** and **h** were analyzed using a two-tailed unpaired non-parametric Mann-Whitney test, data in **d**, **g**, and **i** were analyzed using Fisher's Exact test, and data in **c**, **d** were analyzed using a two-tailed paired non-parametric Wilcoxon matched-pairs signed rank test. See also Supplementary Fig. 1 for gating strategy for panel **h**.

Source data

Extended Data Table 1 Donor information and demographics. Related to Fig. 1

Extended Data Table 2 Anchor epitope-binding mAb information.

Related to Fig. 1 and 2

Extended Data Table 3 Mutation information for Extended Data Fig. 2f,g

Extended Data Table 4 Cryo-EM data collection and refinement statistics for 222-1C06 binding H1. Related to Fig. 2

Supplementary information

Supplementary Information

This file contains Supplementary Fig. 1 and Supplementary Tables 1–6.

Reporting Summary

Peer Review File

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**](#)

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

Guthmiller, J.J., Han, J., Utset, H.A. *et al.* Broadly neutralizing antibodies target a haemagglutinin anchor epitope. *Nature* **602**, 314–320 (2022).
<https://doi.org/10.1038/s41586-021-04356-8>

- Received: 06 August 2021
- Accepted: 15 December 2021
- Published: 23 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04356-8>

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/s41586-021-04356-8>

- Article
- Open Access
- [Published: 22 December 2021](#)

Local and systemic responses to SARS-CoV-2 infection in children and adults

- [Masahiro Yoshida](#) [ORCID: orcid.org/0000-0002-3521-5322](#)^{1,2 na1},
- [Kaylee B. Worlock](#) [ORCID: orcid.org/0000-0002-5656-7634](#)^{1 na1},
- [Ni Huang](#)^{3 na1},
- [Rik G. H. Lindeboom](#) [ORCID: orcid.org/0000-0002-3660-504X](#)^{3 na1},
- [Colin R. Butler](#)^{4,5},
- [Natsuhiko Kumasaka](#) [ORCID: orcid.org/0000-0002-3557-0375](#)³,
- [Cecilia Dominguez Conde](#) [ORCID: orcid.org/0000-0002-8684-4655](#)³,
- [Lira Mamanova](#) [ORCID: orcid.org/0000-0003-1463-8622](#)³,
- [Liam Bolt](#) [ORCID: orcid.org/0000-0001-7293-0774](#)³,
- [Laura Richardson](#) [ORCID: orcid.org/0000-0002-8075-3816](#)³,
- [Krzysztof Polanski](#) [ORCID: orcid.org/0000-0002-2586-9576](#)³,
- [Elo Madissoon](#)^{3,6},
- [Josephine L. Barnes](#) [ORCID: orcid.org/0000-0001-9938-3176](#)¹,
- [Jessica Allen-Hytten](#) [ORCID: orcid.org/0000-0003-4644-0362](#)¹,
- [Eliz Kilich](#)⁷,
- [Brendan C. Jones](#) [ORCID: orcid.org/0000-0002-2637-8859](#)^{4,5},
- [Angus de Wilton](#)⁷,
- [Anna Wilbrey-Clark](#)³,
- [Waradon Sungnak](#) [ORCID: orcid.org/0000-0002-0136-4960](#)³,
- [J. Patrick Pett](#) [ORCID: orcid.org/0000-0002-5249-444X](#)³,
- [Juliane Weller](#) [ORCID: orcid.org/0000-0002-1310-6168](#)³,
- [Elena Prigmore](#) [ORCID: orcid.org/0000-0001-8870-0316](#)³,
- [Henry Yung](#)^{1,7},
- [Puja Mehta](#)^{1,7},
- [Aarash Saleh](#)⁸,
- [Anita Saigal](#)⁸,
- [Vivian Chu](#) [ORCID: orcid.org/0000-0003-1376-1186](#)⁸,
- [Jonathan M. Cohen](#) [ORCID: orcid.org/0000-0003-1004-5598](#)⁷,

- [Clare Cane](#) ORCID: orcid.org/0000-0002-3545-768X⁸,
- [Aikaterini Iordanidou](#)⁸,
- [Soichi Shibuya](#) ORCID: orcid.org/0000-0002-4274-1312⁴,
- [Ann-Kathrin Reuschl](#)⁹,
- [Iván T. Herczeg](#) ORCID: orcid.org/0000-0003-1281-3634¹,
- [A. Christine Argento](#)¹⁰,
- [Richard G. Wunderink](#) ORCID: orcid.org/0000-0002-8527-4195¹⁰,
- [Sean B. Smith](#)¹⁰,
- [Taylor A. Poor](#)¹⁰,
- [Catherine A. Gao](#) ORCID: orcid.org/0000-0001-5576-3943¹⁰,
- [Jane E. Dematte](#)¹⁰,
- [NU SCRIPT Study Investigators](#),
- [Gary Reynolds](#) ORCID: orcid.org/0000-0002-8142-8708¹¹,
- [Muzlifah Haniffa](#) ORCID: orcid.org/0000-0002-3927-2084^{3,11},
- [Georgina S. Bowyer](#) ORCID: orcid.org/0000-0002-2058-4045¹²,
- [Matthew Coates](#)^{12,13},
- [Menna R. Clatworthy](#)^{3,12},
- [Fernando J. Calero-Nieto](#) ORCID: orcid.org/0000-0003-3358-8253¹⁴,
- [Berthold Göttgens](#) ORCID: orcid.org/0000-0001-6302-5705¹⁴,
- [Christopher O'Callaghan](#)^{4,5},
- [Neil J. Sebire](#)^{4,5},
- [Clare Jolly](#) ORCID: orcid.org/0000-0002-4603-2281⁹,
- [Paolo De Coppi](#) ORCID: orcid.org/0000-0002-1659-0207^{4,5},
- [Claire M. Smith](#) ORCID: orcid.org/0000-0002-8913-0009⁴,
- [Alexander V. Misharin](#) ORCID: orcid.org/0000-0003-2879-3789¹⁰,
- [Sam M. Janes](#) ORCID: orcid.org/0000-0002-6634-5939^{1,7},
- [Sarah A. Teichmann](#) ORCID: orcid.org/0000-0002-6294-6366^{3,15},
- [Marko Z. Nikolić](#) ORCID: orcid.org/0000-0001-6304-6848^{1,7 na2} &
- [Kerstin B. Meyer](#) ORCID: orcid.org/0000-0001-5906-1498^{3 na2}

[Nature](#) volume 602, pages 321–327 (2022)

- 55k Accesses
- 754 Altmetric
- [Metrics details](#)

Subjects

- [Infection](#)
- [RNA sequencing](#)
- [SARS-CoV-2](#)
- [Transcriptomics](#)

Abstract

It is not fully understood why COVID-19 is typically milder in children^{1,2,3}. Here, to examine the differences between children and adults in their response to SARS-CoV-2 infection, we analysed paediatric and adult patients with COVID-19 as well as healthy control individuals (total $n = 93$) using single-cell multi-omic profiling of matched nasal, tracheal, bronchial and blood samples. In the airways of healthy paediatric individuals, we observed cells that were already in an interferon-activated state, which after SARS-CoV-2 infection was further induced especially in airway immune cells. We postulate that higher paediatric innate interferon responses restrict viral replication and disease progression. The systemic response in children was characterized by increases in naive lymphocytes and a depletion of natural killer cells, whereas, in adults, cytotoxic T cells and interferon-stimulated subpopulations were significantly increased. We provide evidence that dendritic cells initiate interferon signalling in early infection, and identify epithelial cell states associated with COVID-19 and age. Our matching nasal and blood data show a strong interferon response in the airways with the induction of systemic interferon-stimulated populations, which were substantially reduced in paediatric patients. Together, we provide several mechanisms that explain the milder clinical syndrome observed in children.

[Download PDF](#)

Main

SARS-CoV-2 infection in children presents with milder disease severity compared with infection in adults^{1,2}. The overall risk of severe COVID-19 in children is even lower than originally believed³, with around two deaths per million. The molecular basis of the differences in disease progression between children and adults is not understood and may hold clues for better treatment of severe SARS-CoV-2 infection.

SARS-CoV-2 uses a host cell-surface protein, angiotensin-converting enzyme 2 (ACE2), as a receptor for cellular entry⁴. Studies suggested that *ACE2* expression is both tissue and age dependent^{5,6}, with the highest expression found in nasal epithelium of healthy adults⁷ and comparatively lower expression in paediatric upper⁸ and lower airways^{6,9}. These differences were proposed to contribute to reduced disease severity

in children, although recent studies have found no correlation with age or infection^{[10,11](#)}.

During the initial antiviral immune response, interferon (IFN) is important in inhibiting viral replication, contributing to both innate and cell-intrinsic immunity^{[12,13](#)}. Severe COVID-19 in adults has been linked to an impaired antiviral response in the nasal epithelium and blood^{[14,15,16](#)}, whereas several other studies highlight the contribution of the IFN response to the pathogenesis^{[17,18](#)}.

As the virus spreads, 14% of symptomatic, unvaccinated adults develop progressive respiratory failure displaying a strong inflammatory immune response^{[19](#)}. Single-cell analysis of this response in adults demonstrated the involvement of various immune cell types, including proinflammatory monocytes/macrophages^{[20](#)}, clonally expanded cytotoxic T cells^{[21,22,23](#)} and neutrophils^{[21](#)}. However, the cell-specific immune responses in children have not been comprehensively characterized. Studies comparing bulk RNA-sequencing (RNA-seq) and cytokine profiles between children and adults suggest a more robust immune response, such as increased levels of IFN γ and interleukin-17 (IL-17A) in the plasma^{[24](#)}, and a reduced antibody response and neutralizing activity against SARS-CoV-2 in children^{[25](#)}. The most recent single-cell transcriptional study analysing the upper airways of children with mild COVID-19 revealed that higher expression of pattern recognition receptor pathways was related to a stronger innate immune response^{[11](#)}. However, differences in the coordination of local and systemic immune responses to SARS-CoV-2 between children and adults including patients with severe COVID-19 remain to be elucidated.

To address these questions and identify paediatric-specific responses in COVID-19, we collected matched nasal, tracheal, bronchial and blood samples from healthy individuals and patients with COVID-19 from infancy to adulthood and analysed them using single-cell transcriptomics combined with protein profiling.

Study cohort and experimental overview

Using single-cell RNA-seq and cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), we examined the effects of COVID-19 in children versus adults, comparing the airway and systemic responses. We recruited 19 paediatric and 18 adult patients with COVID-19, ranging from asymptomatic to severe, and 41 healthy children and adults, to profile the cellular landscape in the airways (nasal, tracheal and bronchial brushings) and in matching peripheral blood mononuclear cells (PBMCs) (Fig. [1a](#) and Extended Data Fig. [1a,b](#)). For 6 patients with COVID-19, blood was also taken at hospital discharge. Furthermore, 15 patients contributed nasal

and/or blood samples 3 months after having severe COVID-19. A summary of patient characteristics and metadata is provided in Extended Data Table [1](#).

Fig. 1: Experimental outline and overview of results.

 figure 1



a, Visual overview of the experimental design, numbers of patients, samples taken and single cells sequenced. **b, c**, Uniform manifold approximation and projection (UMAP) visualization of annotated airway epithelial cells (**b**) and immune cells (**c**), with the cell numbers per cell type shown in parentheses. B_{mem}, memory B cells; cDCs, conventional dendritic cells; fDCs, follicular dendritic cells; ILCs, innate lymphoid cells; LCs, Langerhans cells; Mac, macrophages; MAIT, mucosal-associated invariant T cells; Mono, monocytes; NKT, natural killer T cells; pDCs, plasmacytoid dendritic cells; T_{fh}, T follicular helper cells; T_{mem}, memory T cells; T_{reg}, regulatory T cells. A full list of abbreviations is provided in the [Supplementary Note](#). **d**, Airway epithelial cells in the same UMAP as **a** with RNA velocity of major epithelial cell types. **e**, The fraction of SARS-CoV-2 viral unique molecular identifiers (UMI) (where ≥ 10 were detected per donor) relative to total UMI per donor, before filtering out of ambient RNA, in descending order coloured by infection collection interval (days). This was calculated as the days between sample collection and estimated onset of infection, based on the first symptom onset or a positive SARS-CoV-2 RT-qPCR test, whichever was reported first for symptomatic patients, and the latter for asymptomatic patients. **f**, The fraction of airway cells with detected SARS-CoV2 mRNA in each cell type (with immune cells in broad categories) in patients with COVID-19 with detected viral RNA (≥ 5 viral UMI per donor following filtering out ambient RNA). $n = 9$.

In total, we generated a dataset of 659,217 cells (an easy-to-use interactive analysis is provided at <https://www.covid19cellatlas.org/>). We characterized the epithelial and immune cell compartments at a high granularity, identifying 59 cell types and states in airways including previously undescribed ones (Fig. 1b,c and Extended Data Fig. 2a, b) and 34 cell types in blood, mostly based on established markers^{23,26}.

New cell subtypes in airway epithelia

The detailed cell type annotation is described in the [Supplementary Note](#), with marker genes and comparison to existing datasets in Extended Data Figs. 2c and 3a–d. Multiple basal, goblet, ciliated and transit epithelial 1 and 2 (secretory to ciliated) cell types reflect the plasticity of the airway compartment^{26,27,28}, with the main differentiation pathways visualized in Fig. 1d. Notably, transit epithelial 1 cells occur mostly in patients with COVID-19, but also in healthy children (Extended Data Fig. 2a) suggesting a function in development and tissue regeneration. Compared with published adult nasal datasets^{14,28}, we annotated cell types with greater granularity, especially for B and T lymphocytes, and we identified three Hillock-like populations^{14,26,27}. The latter are all marked by *KRT14*, *KRT6A* and *KRT13*, which form a distinct differentiation trajectory (Fig. 1d) similar to the one reported in mice²⁷. Moreover, monocytes fall into clearly distinct clusters, annotated by their highly expressed markers, IL-6⁺ monocytes, GPBAR1⁺ monocytes and CXCL10⁺ monocytes, and were mostly derived from neonates with COVID-19 (Fig. 1c and Extended Data Figs. 2a and 3a).

SARS-CoV-2 reads in airway epithelium

In COVID-19-positive nasal samples, we detected viral reads ($n \geq 10$) in 10 out of 28 patients, with the highest levels found in patients who were sampled closest to the estimated onset of infection (Fig. 1e). After filtering ambient RNA, the cell types with the highest proportion of viral reads were goblet 2 inflammatory cells, followed by cycling basal, transit epithelial and ciliated cells (Fig. 1f), largely mirroring *ACE2* expression (Extended Data Fig. 4). Viral reads were also detected in lymphocytes and myeloid cells (mostly macrophages), reflecting either active infection in macrophages²⁹ or merely uptake of virions or infected cells. The expression of SARS-CoV-2 viral entry and associated factors, including *ACE2*, was similar between children and adults, with few genes correlating with active viral infection (Extended Data Fig. 4a,b). In adults, *ACE2* expression is induced by IFN³⁰ and in response to infection²⁸, but we observed no significant increase of *ACE2* expression in children with COVID-19 (Extended Data Fig. 4c), consistent with recent bulk RNA-seq

comparisons¹⁰. As reported^{31,32}, no SARS-CoV-2 viral reads were detected in peripheral blood.

Airway cell type proportions in COVID-19

We next examined changes in cell type proportions for location, age group and COVID-19 status in all of the airway cell populations (Fig. 2a and Extended Data Fig. 5a, b). To test significance, we used a Poisson linear mixed model ([Methods](#)), enabling us to test the whole cohort in a single analysis while taking into account clinical metadata and technical factors (Extended Data Fig. 5b). Airway epithelial cell type composition showed trends of decreasing basal 1 and increasing secretory and goblet cells with age (Extended Data Fig. 3e), reflecting developmental trajectories from progenitors to differentiated cells (Fig. 1d). Notably, there were significant changes with location, as previously reported³³.

Fig. 2: Differences in airway epithelial and immune cells between paediatric and adult patients with COVID-19.

 figure 2

a, The fold change in and statistical significance of major airway cell type proportions across location of sampling, age group and COVID-19 status, estimated by fitting Poisson generalized linear mixed models taking into account other technical and biological variables ([Methods](#)). The red circles indicate local true sign rate (LTSR) > 0.95. Paed, paediatric. **b**, Comparison of the expression signature of cellular response to IFN α , IFN γ , TNF signalling and neutrophil migration signalling across COVID-19 status and age groups. Neut, neutrophil. **c, d**, Heat maps comparing these expression signatures in healthy paediatric versus adult individuals, and in paediatric versus adult

patients with COVID-19 in epithelial cells (**c**) (the colours indicate difference in scoring) and in immune cells (**d**). **e**, Comparison of expression signatures across COVID-19 severity and age groups. **f**, Representative five enriched Gene Ontology terms in genes that are upregulated in COVID-19 samples in transit epithelial 1 cells, goblet 2 inflammatory cells and IL-6⁺ monocytes. **g**, Immunohistochemistry confocal microscopy image showing S100A9 expression (green) by epithelial cells (EPCAM, magenta) in the nasal epithelium. Nuclei were stained with DAPI (blue). Scale bar, 20 μ m. One representative section out of four technical replicates is shown. For **b**, **e**, pairwise comparisons were performed using two-sided Wilcoxon rank-sum tests; NS, $P > 0.05$; **** $P < 2.2 \times 10^{-16}$.

Contrasting epithelial cells in COVID-19 versus healthy adults, the most highly enriched cell types are transit epithelial 1 and goblet 2 inflammatory cells (Fig. [2a](#); all of the cell types are shown in Extended Data Fig. [5a](#)). We hypothesized that the increased transit epithelial cell numbers reflects a compensatory replacement of dying ciliated cells^{[14,34](#)} by their precursors, to maintain homeostasis after infection as seen in the lower airways^{[35,36](#)}, and consistent with trajectory analysis (Fig. [1d](#)). This is further supported by the return to healthy cell population levels in patients after COVID-19 (Fig. [2a](#)). In adults, the proportions of nasal immune cells were not significantly changed in COVID-19.

In children, epithelial cell proportions did not change but in the immune compartment IL6⁺ monocytes were significantly enriched in COVID-19, with a trend towards higher CXCL10⁺ monocytes and neutrophils. We also observed changes in immune cell populations over healthy childhood (Fig. [2a](#)), such as high monocytes and low CD8 T cell levels in infants, and expansion of B cell populations in young children, reflecting a switch from innate to adaptive immunity^{[37](#)}.

Distinct changes in children and adults

We next examined gene expression changes in children versus adults, in healthy individuals, patients with COVID-19 and patients after COVID-19. In nasal epithelial cells, the biggest changes were observed for gene expression signatures associated with IFN α signalling (Fig. [2b](#)). Healthy adults had the lowest IFN α response that was strongly induced in COVID-19 and returned to preinfection levels in patients after COVID-19. In children, this gene signature was already activated and increased only slightly after infection. These patterns were repeated for signatures of IFN γ response, TNF signalling and neutrophil migration, albeit with smaller fold changes. For nasal immune cells, the induction of the IFN α response signature was higher in children than in adults. The other signatures examined also showed greater induction in children than in adults.

Examining these responses by cell types in healthy children versus adults, the IFN response signatures were already activated in children across several epithelial cell types, with the highest levels in goblet inflammatory cells, Hillock precursors and rare melanocytes (Fig. 2c; absolute values per cell type are shown in Extended Data Fig. 5c). However, SARS-CoV-2-induced IFN responses were higher in adults across many epithelial cell types. Many immune cell types in healthy children had elevated IFN response signatures compared to adults, particularly CD56^{lo} natural killer cells, natural killer T cells, neutrophils, CXCL10⁺ monocytes and some CD8⁺ T cell subsets for IFN α , and a wider range for IFN γ (Fig. 2d). After infection, we observed a greater induction of these responses in immune cells in children, most prominently in monocytes, including in the already expanded IL6⁺ monocytes, CD4 CCR4⁺ T cells and T follicular helper cells.

In adults with COVID-19, a higher systemic IFN response has been reported for non-severe disease^{14,38,39}. We confirmed this across disease severity in our adult cohort for the local response, finding a higher IFN α response in asymptomatic/mild versus moderate/severe disease in both epithelial and immune cells (Fig. 2e). In children, this phenomenon was much stronger in immune versus epithelial cells. These data suggest that, in both children and adults, a strong local IFN response is associated with a milder disease severity, presumably because interferons inhibit viral replication¹³. However, in children, this local response is preactivated in epithelial cells and stronger in immune cells, providing better protection against the virus.

We next examined differential gene expression patterns in healthy versus COVID-19 samples, followed by Gene Ontology term enrichment, in cell types that are particularly associated with disease: transit epithelial 1 and goblet 2 inflammatory cells upregulated in adult COVID-19, and IL-6 monocytes upregulated in children, as strong IFN α responders (Fig. 2f). For transit epithelial cells, this highlighted the IFN type I and II response as well as neutrophil chemotaxis, a notable finding given that neutrophil infiltration is linked to COVID-19 severity⁴⁰. The neutrophil recruitment signature was driven by *S100A8* and *S100A9* expression (calprotectin) (Extended Data Fig. 5d), which is also a key correlate of disease severity⁴¹. For goblet 2 inflammatory cells and IL-6⁺ monocytes, the top two terms were type I IFN signalling and negative viral replication. Enrichment of motile cilium assembly is consistent with our observation that in disease there seems to be a higher cell turnover with precursors such as secretory cells differentiating to replace dying ciliated cells.

As calprotectin expression has primarily been associated with myeloid cells, we validated expression at the protein level in epithelial cells. Figure 2g shows double-positive cells, staining for both calprotectin subunit S100A9 and the epithelial marker EPCAM, in a posterior nasal space biopsy of an adult patient with COVID-19. At the

RNA level, calprotectin is expressed across different secretory cell types (Extended Data Fig. 3b).

Multi-omic blood immune landscape

Using CITE-seq and single-cell profiling of blood from paediatric and adult patients with COVID-19, we annotated 422,220 high-quality single-cell transcriptomes from healthy donors, donors with COVID-19 or donors who had recovered from COVID-19, into 34 blood cell types (Fig. 3a; marker expression and annotation validation is shown in Extended Data Fig. 6a–c). To investigate how the immune system responds to SARS-CoV-2, and how age affects this response, we calculated the fold changes in the proportions of cell types that can be attributed to disease state and age (Fig. 3b and Extended Data Fig. 6d–g). Importantly, our Poisson linear mixed model enabled us to distinguish the immune dynamics that can be attributed to technical effects, ageing and COVID-19. Furthermore, we included an interaction between adulthood and disease status to uncover paediatric-specific immune responses to COVID-19 (Fig. 3b). We observed higher plasma cell and plasmablast proportions, as well as a reduction in the monocyte and dendritic cell compartment in the blood of both adult and paediatric patients with COVID-19, as previously reported in adults^{21,23}.

Fig. 3: Differences in immune response between paediatric and adult patients with COVID-19.

 figure 3

a, UMAP visualization of 422,220 PBMCs incorporating both protein- and RNA-expression data. AS-DC, AXL⁺SIGLEC6⁺ dendritic cells; Baso/Eos, basophil/eosinophil; CM, central memory; CTL, cytotoxic T lymphocytes; EM,

effector memory; EMRA, effector memory re-expressing CD45RA; invar, invariant; n-sw, non-switched; RBCs, red blood cells; sw, switched. **b**, The fold changes in the proportions of immune cell types across age group and disease status, taking into account confounding factors ([Methods](#)). Only cell types that change with a local true sign rate of >0.90 in the disease status groups are shown (all of the cell types are shown in Extended Data Fig. [6d,e](#)). This analysis does not include the cells analysed in **f**. **c**, The fraction of unique TCR sequences in different age groups. **d**, Cell-type-marker expression alongside IFN-stimulated genes. The colour is scaled to all other cell types (Extended Data Fig. [6a](#)). HPC, haematopoietic progenitor cell. **e**, The percentage of IFN-stimulated PBMCs of each symptomatic patient with COVID-19, grouped by the weeks since the onset of symptoms. **f**, Dot plot as in **b** showing the IFN-stimulated subpopulations (IFN-stim) across age and disease status. **g**, Correlation analysis comparing the blood and nose, using a Spearman rank-order correlation coefficient between relative proportion of PBMC subtypes (y axis) and nasal cell types (x axis) (Extended Data Fig. [7d,e](#)) **h**, IFN stimulation in PBMCs and nasal cells, and nasal IFN production in individuals with matched nasal and PBMC data (detailed gene expression dynamics are shown in Extended Data Fig. [8](#)). Dots in **c**, **e** represent independent patient samples. For **e**, the box plots show the median (centre line), the first and third quartiles (box limits), and the whiskers extend to the lowest and highest values within $1.5 \times$ interquartile range. All cell type abbreviations are provided in the [Supplementary Note](#).

Reduced cytotoxic response in children

In contrast to the aforementioned cell types that change consistently in adults and children in response to COVID-19, we observed opposing changes in the abundance of many other immune cell types (Fig. [3b](#)). The circulating immune system of adult patients with COVID-19 is characterized by an increased cytotoxic compartment, in which CD8 $^{+}$ cytotoxic T lymphocytes and effector memory cells re-expressing CD45RA are significantly more abundant in adults. Notably, the latter populations, natural killer cells and CD4 $^{+}$ cytotoxic T lymphocytes are reduced in paediatric patients with COVID-19. Together, this could reflect a more systemic infection and inflammation in adults, whereas the infection in paediatric patients remains more restricted to airways.

Naive T cells in children with COVID-19

In addition to a reduced cytotoxic cellular composition, we observed a striking increase in naive lymphocytes in the blood of paediatric patients with COVID-19 (Fig. [3b](#)). High numbers of naive cells may be attributed to an increased release of immature B and T lymphocytes from the bone marrow and thymus, respectively, or due to

migration of more mature cells to the site of infection. With our statistical model and large cohort of healthy individuals, the strong effects of age on the immune landscape were deconvoluted from the COVID-19 effects into independent age effects and quantified in Fig. 3b. The strong maturation patterns and shift from innate to adaptive immunity observed over healthy childhood amplifies some of the paediatric-specific COVID-19 responses; that is, not only do children have a more naive and reduced cytotoxic response to COVID-19, but they also start off with an immune state that is already skewed towards this response.

Diverse immune repertoire in children

As we detected more naive immune cells in children, we hypothesized that this could affect the amount of unique T and B cell receptors (TCRs and BCRs) that are available to detect new pathogens. Indeed, we observed that the pool of detected TCRs becomes increasingly dominated by expanded clones over age (Fig. 3c and Extended Data Fig. 7a), reducing the amount of unique TCRs that are available to detect unseen pathogens. It is therefore conceivable that a higher TCR repertoire diversity in children could contribute to a faster, more efficient adaptive immune response to SARS-CoV-2.

IFN-stimulated cell subtypes in blood

When annotating our PBMC dataset, we noticed further cell type heterogeneity that generated distinct clusters within all major immune cell types due to high expression of IFN-stimulated genes (Fig. 3d and Extended Data Fig. 7b). Activation of IFN signalling is a key hallmark of COVID-19, acting both as an important protective pathway that can equally be associated with severe COVID-19 (refs. 15,42,43). Although we and others reported an association between global changes in IFN related gene expression and COVID-19 (ref. 23), our increased granularity enabled us to distinguish multiple distinct stimulated and unstimulated populations alongside each other within donors. Importantly, this shows that IFN stimulation of PBMCs does not lead to a global activation of gene expression, but is restricted to a subset of circulating cells.

IFN response in early COVID-19

When investigating the COVID-19 IFN response, we found that IFN-stimulated natural killer, B, T and haematopoietic progenitor cell subpopulations are much more abundant in adult patients compared with paediatric patients with COVID-19 (Fig. 3e, f). In adults, the amount of IFN-stimulated PBMCs is strongly correlated with sampling time since onset of symptoms (Fig. 3e). This suggests that IFN-stimulated

PBMCs are a characteristic of the acute phase of infection, when the innate immune response is trying to control the viral infection. In children, the correlation with onset of symptoms is completely absent (Fig. 3e) but IFN-stimulated cells were abundant in some asymptomatic children (Extended Data Fig. 7c), suggesting a much faster induction and clearance of IFN-stimulated cells. Together, these observations support our hypothesis that COVID-19-induced inflammation and cytotoxicity in the blood is more abundant in adults than in children.

Dendritic cells initiate IFN response

To investigate the connection between the local and systemic immune response to SARS-CoV-2, we compared cell type proportions in the blood and nose for multi-tissue donors and observed strong correlations (Fig. 3g; all comparisons are shown in Extended Data Fig. 7d,e). Particularly, SARS-CoV-2-infected and inflammatory nasal epithelial cells, and nasal plasmacytoid and conventional dendritic cells correlated with IFN stimulation in the blood. This is interesting as dendritic cells are known for their viral-sensing and IFN-production capacities⁴⁴, but this has not been directly observed in COVID-19. Although dendritic cells protect against severe disease⁴⁵, most COVID-19 studies that analysed blood reported a depletion of dendritic cells⁴⁶. However, here we provide evidence that, at the earliest stages of infection, type I and type III IFNs are detectable (Fig. 3h) and are produced by plasmacytoid and conventional dendritic cells, but not other immune or epithelial cells (Extended Data Fig. 8b and [Supplementary Note](#)).

Discussion

Here we focused on why children are generally protected from severe COVID-19 and propose multiple mechanisms (Fig. 4). First, we show that the airway epithelium has a higher steady-state expression of IFN-response genes in children. SARS-CoV-2 has been reported to be highly sensitive to prestimulation with interferons⁴⁷, and preactivation may restrict viral spread in children. Second, the systemic immune response in blood is characterized by a more naive state. By contrast, adults display a highly cytotoxic immune compartment in the blood, probably due to a failure to restrict viral spreading. This elevated systemic response in adults can lead to widespread immune-related organ damage⁴⁸. A third feature that we observed was the higher TCR repertoire diversity in children versus adults. The acquisition of memory T and B cells during childhood and adulthood, combined with reduced thymic output, shifts the adaptive immune system into a more memory-based compartment in aged individuals⁴⁹. This reduces the pool of unique immune receptors within naive lymphocytes⁵⁰, making it less probable that a high-affinity immune receptor is directly available against SARS-CoV-2 antigens. Finally, we found previously undescribed

IFN-stimulated cell states in multiple blood cell lineages that are highly abundant in early disease in adults. This presents an added inflammatory feature of the already cytotoxic immune compartment in adult patients with COVID-19, and possibly amplifies any pathological effects of the systemic immune response. The identification of both IFN-stimulated and unstimulated blood cells within donors underscores that activation is cell specific rather than, as noted by others, systemic, possibly caused by either close proximity to the site of infection or an associated secondary lymphoid organ, or cell-to-cell variability in responsiveness as we have shown in fibroblasts and phagocytes⁵¹.

Fig. 4: The local and systemic response to SARS-CoV-2 infection in children and adults.

 figure 4

Schematic of the difference in the airway and systemic immune response to SARS-CoV-2 infection between children and adults, reflecting the maturation of the immune landscape throughout childhood to adulthood. The key points are (1) immune cell proportions display strong maturation patterns throughout healthy childhood and adulthood, with a notable innate to adaptive immunity switch. (2) In the airways, the local innate IFN response to SARS-CoV-2 is stronger in paediatric airway immune cells compared with adult airway immune cells. (3) In the blood, the systemic innate IFN response to SARS-CoV-2 is stronger in adults, with a notable increase in IFN-stimulated subpopulations, whereas the adaptive immune response is characterized by expanded cytotoxic populations in adults compared with naive populations in children. (4) Epithelial cells with an inflammatory gene expression (S100A8/S100A9) are found

enriched in patients with COVID-19. (5) Clonotype diversity decreases with age. The figure was generated using BioRender.com.

SARS-CoV-2 infection frequently starts in the upper airways, in which we found the highest total viral load in surface epithelial goblet, ciliated and differentiating cells. Viral infections are cleared by cell death and the removal of the infected cells⁵², leading to a highly dynamic restructuring of the airway epithelium with a marked increase in developmental intermediates, most notably the transit epithelial populations, which are rebalanced after infection. We also observed a strong neutrophil recruiting signature, driven by the expression of calprotectin in epithelial cell types, highlighting the key role of epithelial cells in initiating an innate immune response.

Overall, our study provides multiple insights using paired multi-omics profiling of both airway epithelium and peripheral blood to fill the gap in our understanding of paediatric epithelial and immune responses to COVID-19, while also identifying previously undescribed cell states in both airway epithelium and blood. These insights could contribute to pinpointing the triggers of severe disease in adults with a view towards risk stratification and therapeutic intervention.

Methods

Study participants and design

The UK cohort

Participants were included from five large hospital sites in London, UK, namely Great Ormond Street Hospital NHS Foundation Trust, University College London Hospitals NHS Foundation Trust, Royal Free London NHS Foundation Trust (Royal Free Hospital and Barnet Hospital) and Whittington Health NHS Trust from March 2020 to February 2021. Ethical approval was given through the Living Airway Biobank, administered through the UCL Great Ormond Street Institute of Child Health (REC reference: 19/NW/0171, IRAS project ID: 261511, North West Liverpool East Research Ethics Committee), REC reference 18/SC/0514 (IRAS project: 245471, South Central Hampshire B Research Ethics Committee) administered through the University College London Hospitals NHS Foundation Trust and REC reference 18/EE/0150 (IRAS project ID: 236570, East of England Cambridge Central Research Ethics Committee) administered through Great Ormond Street Hospital NHS Foundation Trust, REC reference 08/H0308/267 administered through the Cambridge University Hospitals NHS Foundation Trust, as well as by the local R&D departments at all hospitals. All of the study participants or their surrogates provided informed

consent. At daily virtual COVID-19 coordination meetings, suitable patients were chosen from a list of newly diagnosed patients who were admitted within the preceding 24 h. Only patients with COVID-19 who tested positive for SARS-CoV-2 by a quantitative PCR with reverse transcription (RT–qPCR) nasopharyngeal test were enrolled in the study; a summary of symptom onset relative to RT–qPCR testing and sampling is shown in Extended Data Fig. 1b. Patients with typical clinical and radiological COVID-19 features but with a negative screening test for SARS-CoV-2 were excluded. Other excluding criteria included active haematological malignancy or cancer, known immunodeficiencies, sepsis from any cause and blood transfusion within 4 weeks. Two cases of paediatric multisystem inflammatory syndrome (PIMS-TS, named by the Royal College of Paediatrics and Child Health) were included (airway samples only), which is also referred to as multisystem inflammatory syndrome in children (MIS-C) by the World Health Organization, with little to no MIS-C-specific difference detected after analysis in the nasal mucosa compared with equivalent samples from paediatric patients with COVID-19 (ref. 53). Maximal severity of COVID-19 was determined retrospectively by determining the presence of symptoms, the need for oxygen supplementation and the level of respiratory support (mild, symptomatic without oxygen requirement or respiratory support; moderate, requiring oxygen without respiratory support; severe, requiring non-invasive or invasive ventilation). Brushings and peripheral blood sampling were performed by trained clinicians before inclusion in any pharmacological interventional trials, with the exception of 3 paediatric patients with COVID-19 (noted in Extended Data Table 1) and ideally within 48 h of a positive SARS-CoV-2 test. All of the participants for our paediatric healthy cohort were recruited from Great Ormond Street Hospital NHS Foundation Trust and were eligible for inclusion if they were <18 years old and asymptomatic for respiratory viral infections at time of sampling. At the start of the study, initiated in March 2020 it was not standard practice for hospitals to test healthy asymptomatic patients. Therefore 8 (out of 30) of the earliest recruited participants were untested and assumed negative. To confirm this assumption and to look for any other undetected asymptomatic infections metagenomic analysis on the entire dataset was performed (see ‘Metagenomic analysis’; Extended Data Fig. 9). Participants for our adult healthy cohort were recruited from University College London Hospitals and associated research laboratories at University College London and were eligible for inclusion if >18 years and asymptomatic with a current negative SARS-CoV-2 test (RT–qPCR or rapid-antigen testing). Exclusion criteria for the cohort included active haematological malignancies or cancer, known immunodeficiencies, sepsis from any cause and blood transfusions within 4 weeks, known bronchial asthma, hay fever, diabetes and other known chronic respiratory diseases such as cystic fibrosis, interstitial lung disease and chronic obstructive pulmonary disease. There were three exceptions to these criteria in our paediatric cohort; NP28 who was later discovered to have asthma and NP10 who was reported to have a immunocompromised status in underlying comorbidities, but for whom only nasal brushes were included; and NP27,

who did not have any respiratory problem but was subsequently diagnosed with endocarditis. Exclusion of these individuals did not alter any of our conclusions. For some patients included in our COVID-19 cohort, matched convalescent blood was taken on the day of hospital discharge and analysed separately in our post-COVID-19 cohort alongside symptomatic patients recruited from University College London Hospitals outpatient COVID-19 follow-up clinic, who were recalled around 3 months after recovering from severe COVID-19 using the exclusion criteria as stated for our COVID-19 cohort. Participants were further divided into subgroups to enable us to look at age-specific effects. These were classified based on the World Health Organisation; neonates (0–30 days), infants (1–24 months), young children (2–5 years), children (6–11 years), adolescents (12–17 years) and adults (≥ 18 years); adults were further broken down into adults (18–64 years) and elderly (≥ 65 years).

Chicago cohort (adult bronchial samples)

Ethical approval for sample collection from patients with severe pneumonia was given by Northwestern Institutional Review Board, study STU00204868 (PI Richard Wunderink). Samples from patients with COVID-19, viral pneumonia and other pneumonia, and non-pneumonia controls were collected from participants enrolled in the Successful Clinical Response in Pneumonia Therapy (SCRIPT) study STU00204868 and admitted to the ICU at Northwestern Memorial Hospital, Chicago. All of the study participants or their surrogates provided informed consent. Individuals aged at least 18 years with suspicion of pneumonia based on clinical criteria (including but not limited to fever, radiographic infiltrate and respiratory secretions) were screened for enrolment into the SCRIPT study. The inability to safely perform bronchoalveolar lavage or non-bronchoscopic bronchoalveolar lavage was considered to be an exclusion criteria. In our centre, patients with respiratory failure are intubated on the basis of the judgement of bedside clinicians for worsening hypoxaemia, hypercapnia or work of breathing refractory to high-flow oxygen or non-invasive ventilation modes. Bronchial brushings were performed during the diagnostic bronchoalveolar lavage procedure and the samples were collected from representative sites at the lobar bronchi.

Sample collection

The UK cohort

Samples were collected and transferred to a category level 3 facility at University College London and processed within 2 h of sample collection. Nasal, tracheal and bronchial brushings were enzymatically digested to a single-cell suspension and processed further immediately. Peripheral blood was centrifuged after adding Ficoll

Paque Plus and PBMCs, serum and neutrophils were separated, collected and frozen for later processing. A local anaesthetic endoscopically guided biopsy of the postnasal space mucosa was collected from a 19-year-old female patient three weeks after onset of mild COVID-19 symptoms (REC: 08/H0308/267). SARS-CoV-2 virus was confirmed by RT–PCR testing at the time of symptom onset.

Chicago cohort (adult bronchial samples)

Samples were collected in the ICU at Northwestern Memorial Hospital, transferred to a research laboratory in the Simpson Querrey Biomedical Research Center, Feinberg School of Medicine, Northwestern University, and processed within 1 h of sample collection in biological safety level 2 facility using biological safety level 3 practices. After collection, bronchial brushings were stored in Hypothermosol (Stem Cell Technologies, 07935) at 4 °C.

Nasal and tracheal brushing tissue dissociation

The UK cohort

Nasal brushing was performed on the inferior nasal concha zone with a cytological brush (Scientific Laboratory Supplies, CYT1050). All of the samples were processed fresh according to a previously described protocol²⁶ with minor modifications⁵⁴. The brushes were immediately placed in a 15-ml sterile Falcon tube containing 4 ml of transport medium (αMEM supplemented with 1× penicillin–streptomycin (Gibco, 15070), 10 ng ml⁻¹ gentamicin (Gibco, 15710) and 250 ng ml⁻¹ amphotericin B (Thermo Fisher Scientific; 10746254)) on ice. Once in the category level 3 facility, the tube was shaken vigorously to collect cells in suspension. The brushes were then carefully transferred into a new Falcon tube containing HBSS and shaken to remove residual cells from the brush. This was repeated until all of the cells looked like they had been collected from the brush. All of the Falcon tubes were centrifuged at 400g for 5 min at 4 °C. The cell pellet was collected from each tube and then put in a dissociation buffer consisting of 10 mg ml⁻¹ protease from *Bacillus licheniformis* (Sigma-Aldrich, P5380) and 0.5 mM EDTA in HypoThermosol (Stem Cell Technologies, 07935) for dissociation on ice for 30 min. Every 5 min, cells were gently triturated using a 21 G and 23 G needle. After incubation, protease was inactivated by adding 200 µl of inactivation buffer (HBSS containing 2% BSA). The suspension was centrifuged at 400g for 5 min at 4 °C and the supernatant was discarded. Cells were resuspended in 1 ml wash buffer (HBSS containing 1% BSA) and centrifuged again. Red blood cell lysis was performed if needed, followed by an additional wash. The single-cell suspension was forced through a 40-µm Flowmi Cell Strainer. Finally, the cells were centrifuged and resuspended in 30 µl of resuspension

buffer (HBSS containing 0.05% BSA). Using Trypan Blue, total cell counts and viability were assessed. The cell concentration was adjusted for 5,000 targeted cell recovery according to the 10x Chromium manual before loading onto the 10x chip (between 700–1,000 cells per μ l) and processing immediately for 10x 5' single-cell capture using the Chromium Single Cell V(D)J Reagent Kits V1.0 (Rev J Guide), the newer chromium Next GEM Single Cell V(D)J Reagent Kit v1.1 (Rev E Guide) or the chromium Next GEM Single Cell 5' V2 (Dual index) kit (Rev A guide).

For a small subset of nasal samples (PP5_NB_2, PP6_NB_2, AP11_NB, AP12_NB, AP13_NB and AP14_NB_2) 1 μ l viral RT oligo (at 5 μ M, PAGE) was spiked into the master mix (at step 1.2.b in the 10x guide; giving a final volume of 75 μ l) to help with the detection of SARS-CoV-2 viral reads. The samples were then processed according to the manufacturer's instructions, with the viral cDNA separated from the gene expression libraries (GEX) by size selection during step 3.2. Here the supernatant was collected (159 μ l) and transferred to a new PCR tube and incubated with 70 μ l of SPRI beads (0.6 \times selection) at room temperature for 5 min. The SPRI beads were then washed according to the guide and the viral cDNA was eluted using 30 μ l of EB buffer. No changes to the transcriptome were observed between samples, which were run both with and without the viral oligo and only a small increase in the overall number of SARS-CoV-2 reads detected. The RT oligo sequence was as follows: 5'-AAGCAGTGGTATCAACGCAGAGTACTCGTGTCCCTGTCAACG-3'

Chicago cohort (adult bronchial samples)

Samples were processed using a previously reported protocol²⁶ with minimal modifications. Specifically, dissociation was performed without EDTA and trituration was performed by pipetting using a regular-bore 1,000 μ l tip every 5 min. Dissociation was visually confirmed by inspecting an aliquot of the single-cell suspension using phase contrast on an inverted microscope. Cell count was performed using the AO/PI reagent on the K2 Cellometer (Nexcelom). Approximately 300,000–500,000 cells were obtained per brush with a viability of 97% and above. Cells were captured on a 10x Chromium Single Cell Controller using the Chromium Single Cell V(D)J Reagent Kits V1.0 (Rev J Guide).

PBMC isolation from peripheral blood

Peripheral blood was collected in EDTA immediately after the nasal brushing procedure. The blood was diluted with 5 ml of PBS containing 2 mM EDTA (Invitrogen, 1555785-038). Diluted blood (10–20 ml) was carefully layered onto 15 ml of Ficoll-Paque Plus (GE healthcare, 17144002). If the sample volume was less than 5 ml, blood was diluted with an equal volume of PBS-EDTA and layered onto 3 ml Ficoll. The sample was centrifuged at 800g for 20 min at room temperature. The

plasma layer was carefully removed and the PBMC layer was collected using a sterile Pasteur pipette. The PBMC layer was washed with 3 volumes of PBS containing EDTA by centrifugation at 500g for 10 min. The pellet was suspended in PBS-EDTA and centrifuged again at 300g for 5 min. The PBMC pellet was collected, and then both cell number and viability were assessed using Trypan Blue. Cell freezing medium (90% FBS, 10% DMSO) was added dropwise to PBMCs slowly on ice and the mixture was then cryopreserved at -80 °C until further full sample processing.

CITE-seq staining for single-cell proteogenomics

Frozen PBMC samples were thawed quickly at 37 °C in a water bath. Warm RPMI1640 medium (20–30 ml) containing 10% FBS was added slowly to the cells before centrifuging at 300g for 5 min. This was followed by a wash in 5 ml RPMI1640-FBS. The PBMC pellet was collected, and the cell number and viability were determined using Trypan Blue. PBMCs from four different donors were then pooled together at equal numbers: 1.25×10^5 PBMCs from each donor were combined with the other PBMCs to make up 5.0×10^5 cells in total. The remaining cells were used for DNA extraction (Qiagen, 69504). The pooled PBMCs were resuspended in 25 µl of cell staining buffer (BioLegend, 420201) and blocked by incubation for 10 min on ice with 2.5 µl Human TruStain FcX block (BioLegend, 422301). The PBMC pool was then stained with TotalSeq-C antibodies (BioLegend, 99814) according to the manufacturer's instructions. For a full list of TotalSeq-C antibodies, refer to ref. [23](#). After incubating with 0.5 vials of TotalSeq-C for 30 min at 4 °C, PBMCs were washed three times by centrifugation at 500g for 5 min at 4 °C. PBMCs were counted again and processed immediately for 10x 5' single cell capture (Chromium Next GEM Single Cell V(D)J Reagent Kit v1.1 with Feature Barcoding technology for cell Surface Protein-Rev D protocol). Two lanes of 25,000 cells were loaded per pool onto a 10x chip.

Library generation and sequencing

The Chromium Single Cell 5' V(D)J Reagent Kit (V1.0 chemistry), Chromium Next GEM Single Cell 5' V(D)J Reagent Kit (V1.1 chemistry) or Chromium Next GEM Single Cell 5' v2 kit (V2.0 chemistry) was used for single-cell RNA-seq library construction for all airway samples, and the Chromium Next GEM Single Cell V(D)J Reagent Kit v1.1 with Feature Barcoding technology for cell surface proteins was used for PBMCs. GEX and V(D)J libraries were prepared according to the manufacturer's protocol (10x Genomics) using individual Chromium i7 Sample Indices. The cell surface protein libraries were created according to the manufacturer's protocol with slight modifications that included doubling the SI primer amount per reaction and reducing the number of amplification cycles to 7 during the index PCR to avoid the daisy chains effect. GEX, V(D)J and cell surface protein indexed libraries were pooled

at a ratio of 1:0.1:0.4 and sequenced on a NovaSeq 6000 S4 Flowcell (paired-end, 150 bp reads) aiming for a minimum of 50,000 paired-end reads per cell for GEX libraries and 5,000 paired-end reads per cell for V(D)J and cell surface protein libraries.

Single-cell RNA-seq computational pipelines, processing and analysis

The single-cell data were mapped to a GRCh38 ENSEMBL 93 derived reference, concatenated with 21 viral genomes (featuring SARS-CoV-2), of which the NCBI reference sequence IDs are: [NC_007605.1](#) (EBV1), [NC_009334.1](#) (EBV2), [AF156963](#) (ERVWE1), [AY101582](#) (ERVWE1), [AY101583](#) (ERVWE1), [AY101584](#) (ERVWE1), [AY101585](#) (ERVWE1), [AF072498](#) (HERV-W), [AF127228](#) (HERV-W), [AF127229](#) (HERV-W), [AF331500](#) (HERV-W), [NC_001664.4](#) (HHV-6A), [NC_000898.1](#) (HHV-6B), [NC_001806.2](#) (herpes simplex virus 1), [NC_001798.2](#) (herpes simplex virus 2), [NC_001498.1](#) (measles morbillivirus), [NC_002200.1](#) (mumps rubulavirus), [NC_001545.2](#) (rubella), [NC_001348.1](#) (varicella zoster virus), [NC_006273.2](#) (cytomegalovirus) and [NC_045512.2](#) (SARS-CoV-2). When examining viral load per cell type, we first removed ambient RNA by SoupX and only included SARS-CoV-2-positive donors where ≥ 5 viral reads were still detected. Antibody-derived tag counts and gene expression counts in CITE-seq data were jointly quantified using Cell Ranger v.3.0.2. The alignment, quantification and preliminary cell calling of airway samples were performed using the STARsolo functionality of STAR v.2.7.3a, with the cell calling subsequently refined using the Cell Ranger v.3.0.2 version of EmptyDrops⁵⁵. This algorithm has been made available as emptydrops on PyPi. Initial doublets were called on a per-sample basis by computing Scrublet⁵⁶ scores for each cell, propagating them through an over-clustered manifold by replacing individual scores with per-cluster medians, and identifying statistically significant values from the resulting distribution, replicating the approach of refs. ^{57,58}. The clustering was performed with the Leiden⁵⁹ algorithm on a k -nearest neighbour graph of a principal component analysis (PCA) space derived from a log[counts per million/100 + 1] representation of highly variable genes, according to the SCANPY protocol⁶⁰, and overclustering was achieved by performing an additional clustering of each resulting cluster. The primary clustering also served as an input for ambient RNA removal using SoupX⁶¹.

Metagenomics analysis

To ensure that the patients in our cohort did not carry undiagnosed infections, we carried out a metagenomic analysis using mg2sc (<https://github.com/julianeweller/mg2sc>). The metagenomic tool Kraken 2 (ref. ⁶²) was installed according to the standard instructions on GitHub^{63,64}. The prebuilt standard Kraken 2 database was downloaded from <https://benlangmead.github.io/aws->

[indexes/k2](#) (standard from 12 February 2020, 36 GB). Only reads that were not aligned to *Homo sapiens* with STARSolo⁶⁵ were extracted from the STARSolo and converted into FASTQ using bedtools (v.2.30)^{66,67} for subsequent metagenomic analysis. This was performed using python scripts available on GitHub (<https://github.com/julianeweller/mg2sc>) and the command ‘scMeG-kraken.py --input [bamfile, e.g. starsolo/Aligned.sortedByCoord.out.bam] --outdir [output directory] / --DBpath [path to kraken database] --threads [#, e.g. 8] --prefix [preferred file prefix] --verbosity [error/warning/info/debug]’ resulting in a matrix of cell barcodes with assigned taxonomy transcript counts. Organisms shown are highly variable between samples with min_mean = 0.08, max_mean = 10 and min_disp = 0.05. The results are shown in Extended Data Fig. 9.

Confocal microscopy method

Nasal epithelial biopsies were placed in Antigenfix (Microm Microtech) for 1–2 h at 4 °C, then 30% sucrose in PBS for 12–24 h at 4 °C, before cryopreservation in OCT (Cell Path). Sections (30 µm) were permeabilized and blocked in PBS containing 0.3% Triton X-100 (Sigma-Aldrich), 1% normal goat serum, 1% normal donkey serum and 1% BSA (R&D) for 1–2 h at room temperature. The samples were stained with anti-human S100A9 antibodies conjugated to FITC (1 in 50 dilution, MRP 1H9, BioLegend) and anti-human EPCAM antibodies conjugated to APC (1 in 50 dilution, MRP14, BioLegend, 350703) in blocking buffer overnight and washed three times for 10 min in PBS before mounting with Fluoromount-G containing DAPI (Invitrogen). Images were acquired using a Leica SP8 confocal microscope. Raw imaging data were processed using Imaris (Bitplane).

Airway single-cell RNA-seq data processing

Quality control, normalization and clustering

To account for large quality variance across different samples, quality control was performed on SoupX-cleaned expression matrixes for each sample separately. Quality control thresholds were automatically established by fitting a 10-component Gaussian mixture model to the log-transformed UMI count per cell and to the percentage of mitochondrial gene expression and finding the lower or higher bounds where the probability density falls under 0.05. We also excluded cells with haemoglobin expression >0.1% of total expression and genes expressed in fewer than 3 cells. Expression values were then normalized to a sum of 1×10^4 per cell and log-transformed with an added pseudocount of 1. Highly variable genes were selected within each sample and then merged with the top 3,000 most commonly found genes chosen using the Scanpy⁶⁰ function scanpy.pp.highly_variable_genes(). After

removing mitochondrial and ribosomal genes from the list of highly variable genes, principle component analysis was performed and the top 30 principle components were selected as input for BBKNN⁶⁸ to correct for batch effects between donors and compute a batch-corrected k -nearest neighbour graph. Leiden clustering was performed on this graph with a resolution of 0.2 to separate broad cell types (epithelial cells, B/plasma cells, T/natural killer/innate lymphoid cells and myeloid cells). For each broad cell type, clustering was then repeated, starting from highly variable gene discovery to achieve a higher resolution and a more accurate separation of refined cell types. Subclusters were manually examined and further reclustered when necessary.

Quantifying SARS-CoV-2 viral expression

For donor-level quantification, we took the data before ambient RNA removal by SoupX, as ambient viral RNA still reflects total viral load. For cell-type-level quantification, we used the data after ambient RNA removal, as ambient viral RNA cannot be assigned to specific cells.

Developmental trajectory inference

RNA velocity analysis was performed to infer developmental trajectory for the major epithelial cell types (excluding melanocytes, ionocytes, brush cells and neuroendocrine cells). Spliced and unspliced UMI counts were generated using the STARSolo functionality of STAR v.2.7.3a. scvelo was used to fit a dynamical model as previously described⁶⁹ on the basis of the top 2,000 highly variable genes with at least 20 UMI for both spliced and unspliced transcripts across all cells.

Expression signature analysis

The gene sets GOBP_response_to_interferon_alpha, GOBP_response_to_interferon_gamma, GOBP_response_to_tumor_necrosis_factor and GOBP_neutrophil_migration were retrieved from the Molecular Signature Database (<http://www.gsea-msigdb.org>)⁷⁰ and the Scanpy function scanpy.tl.score_genes() was used to score the signature for each cell.

CITE-seq data processing

Demultiplexing and doublet removal of PBMC samples

For pooled donor CITE-seq samples, the donor ID of each cell was determined by genotype-based demultiplexing using souporcell (v.2)⁷¹. Souporcell analyses were

performed with skip_remap enabled and a set of known donor genotypes given under the common_variants parameter. The donor ID of each souporcell genotype cluster was annotated by comparing each souporcell genotype to the set of known genotypes. Droplets that contained more than one genotype according to souporcell were flagged as ‘ground-truth’ doublets for heterotypic doublet identification. Ground-truth doublets were used by DoubletFinder (v.2.0.3)^{[72](#)} to empirically determine an optimal pK value for doublet detection. DoubletFinder analysis was performed on each sample separately using 10 principal components, a pN value of 0.25, and the nExp parameter estimated from the fraction of ground-truth doublets and the number of pooled donors.

CITE-seq background and ambient RNA subtraction

Background and non-specific staining by the antibodies used in CITE-seq was estimated using SoupX (v.1.4.8)^{[61](#)}, which models the background signal on near-empty droplets. The soupQuantile and tfidfMin parameters were set to 0.25 and 0.2, respectively, and lowered by decrements of 0.05 until the contamination fraction was calculated using the autoEstCont function. Gene expression data were also corrected using SoupX to remove cell-free mRNA contamination using the default SoupX parameters.

CITE-seq quality control and normalization

CITE-seq data were filtered by removing droplets with fewer than 200 genes expressed or with more than 10% of the counts originating from mitochondrial genes. Gene expression data were normalized with a log + 1 transformation (log1p), and 2,000 hyper variable genes were selected using the vst algorithm in Seurat (v.3.9.9.9024)^{[73](#)}. Antibody-derived tag counts were normalized with the centred log-ratio transformation.

Integrated embedding and clustering of CITE-seq data

PCA was run separately on gene expression and antibody-derived tag count data, followed by batch correction using harmony⁷⁴ on the sequencing library identifier. Nearest neighbour graphs and UMAP visualizations were generated based on the first 30 harmony-adjusted principal components. The first 30 harmony-adjusted principal components of both gene expression and antibody-derived tag count data were used to compute a weighted nearest neighbour graph⁷⁵ with Seurat and embedded using UMAP. Cells were clustered with the Leiden algorithm using the igraph R package, with a resolution of 4. After initial clustering of all PBMCs, subsets of all T and natural killer cells, all B and plasma cells, and all monocytes and dendritic cells were reclustered after hypervariable gene selection within each subset. Cells in weighted-nearest-neighbour-based clusters with less than 100 members were reassigned on the basis of the closest multimodal neighbour.

Comparison of PBMCs using Azimuth

The manual blood cell type annotation was validated using the Azimuth tool (<https://azimuth.hubmapconsortium.org>). A randomly sampled subset of 100,000 PBMCs were uploaded to predict their cell type identity.

Differential expression analysis in airway data

In addition to the differential expression analysis, correcting for various metadata, that was performed on the whole-airway and PBMC datasets as described below, results shown for subsets of the data were obtained with a simpler method. After subsetting cell types and/or age groups, a Wilcoxon rank-sum test (implemented in Scanpy⁶⁰) was performed to compare gene groups. The sets of differentially expressed genes were further analysed using the g:Profiler toolkit⁷⁶ (g:Profiler version e102_eg49_p15_7a9b4d6, database updated on 15 December 2020) for functional enrichment analysis. The expression of SARS-CoV-2 viral entry factors, including ACE2 and secondary entry receptors (*NRPI* (refs. ^{77,78}), *BSG*⁷⁹, *TFRC*⁸⁰), along with

other viral-entry-associated factors, were analysed in each cell type (Extended Data Fig. 4a).

Defining the interferon-stimulated signature in blood

The genes that make up the interferon-stimulated signature in blood were defined by performing Wilcoxon rank-sum tests in Seurat between each interferon-stimulated subpopulation and its matched unstimulated population. The genes that were most significant (false-discovery rate not distinguishable from 0) in all comparisons were included in the interferon-stimulated signature shown in Fig. 3. This list includes *BST2*, *CMPK2*, *EIF2AK2*, *EPSTI1*, *HERC5*, *IFI35*, *IFI44L*, *IFI6*, *IFIT3*, *ISG15*, *LY6E*, *MX1*, *MX2*, *OAS1*, *OAS2*, *PARP9*, *PLSCR1*, *SAMD9*, *SAMD9L*, *SP110*, *STAT1*, *TRIM22*, *UBE2L6*, *XAF1* and *IRF7*.

Inference of ethnicity from single-cell RNA-seq data

The latest biallelic single-nucleotide polymorphism (SNP) genotype data (GRCh38) was obtained from the 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/). Allele-specific counts of RNA-seq reads at the SNP location in 1000 Genomes Project data were generated for each airway sample. As the read coverage from the single-cell RNA-seq data was strongly enriched around the 5' end of a gene, SNP loci covered at least 20 reads for more than 90% of samples that were used (19,733 genome-wide SNP loci in total). The SNP genotype from allele-specific expression was determined as a maximum posterior genotype after fitting a beta-binomial mixture distribution with underlying probabilities of 0.01, 0.5 and 0.99 for reference homozygote, heterozygote and alternative homozygote, respectively. The overdispersion parameter of the beta-binomial distribution was estimated for each sample independently shared across all SNPs. The genotype data from 1000 Genomes samples were combined with the genotype data for our samples, and PCA was performed on the scaled genotype data (mean 0 and s.d. equal to 1 for each SNP locus). The ethnicity of each sample was determined by the Mahalanobis distance to the four major ethnic groups in the 1000 Genomes Project

(African, East Asian, European and South Asian). The first three principal components were used to compute the cluster centre and the covariance matrix for each ethnic group.

Cell type composition analysis

The number of cells for each sample and cell type combination was modelled with a generalized linear mixed model with a Poisson outcome. The five clinical factors (age, sex, inferred ethnicity, tissue and the interaction of COVID-19 status and broad age group) and three technical factors (donor, 10x kit, sequencing batch and sample) were fitted as random effects to overcome the collinearity among the factors. The effect of each clinical/technical factor on cell type composition was estimated by the interaction term with the cell type. The glmer function in the lme4 package implemented on R was used to fit the model. The standard error of the variance parameter for each factor was estimated using the numDeriv package. The conditional distribution of the fold change estimate of a level of each factor was obtained using the ranef function in the lme4 package. The log-transformed fold change is relative to the grand mean and adjusted such that it becomes 0 when there is no effect. The statistical significance of the fold change estimate was measured by the local true sign rate (LTSR), which is the probability that the estimated direction of the effect is true, that is, the probability that the true log-transformed fold change is greater than 0 if the estimated mean is positive (or less than 0 if the estimated mean is negative). It is calculated on the basis of the estimated mean and s.d. of the distribution of the effect (log-transformed fold change), which is to an extent similar to performing a (one-sided) one-sample Z-test and showing $(1 - P)$.

Differential expression analysis using metadata

We performed differential gene expression analysis for both airway and PBMC data. We used the 7 clinical (donor, age group, sex, ethnicity, tissue, smoking status and COVID-19 status) and the 4 technical factors (batch, 10x kit version, the number of expressed genes and the number of mapped fragments) to adjust for confounding effects. For PBMC data, the tissue and

10x kit were identical across samples and not included in the model. We used the linear mixed model proposed in ref. ⁸¹ to adjust for the 11 confounding factor effects and the effect of cell type as a random effect in differential expression analysis. We fit the model on a gene-by-gene basis using the estimated variance parameters to test each factor k explaining a significant amount of transcription variation. If the focal factor k is a categorical variable with L levels (for example, COVID-19 status with 3 levels), we partitioned the levels into one of two groups. There are $2L - 1$ contrasts that were tested against the null model (removing the focal factor k in the model) to compute Bayes factors. Those Bayes factors were next used for fitting a finite mixture model to compute the posterior probability as well as the LTSR (see section 1.3 of the supplementary note of ref. ⁸¹ for more details). We used g:Profiler 2 implemented in R (v.2.0.1.5) to identify which pathways are enriched for differentially expressed genes for each contrast. We used genes of which the LTSR is greater than 0.5 to perform the analysis (both upregulated and downregulated genes separately).

Single-cell VDJ-sequencing data analysis

TCR and BCR sequencing data were processed using Cell Ranger and downstream analysis was performed using the scirpy package (v.0.6.1)⁸². In brief, we integrated TCR and BCR data with gene expression from T cell and B cell subsets, respectively. After categorizing cells on the basis of the detection of productive antigen receptor chains, we selected cells with a single pair of productive chains for further analysis. T cell clonotypes were defined at the amino acid level, considering both receptor chains. B cell clonotypes were defined at the amino acid level while allowing for a Hamming distance of up to 10% of the sequence, considering both receptor chains.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The dataset from our study can be explored interactively through a web portal (<https://covid19cellatlas.org>). Quality control metrics for our single-cell data are provided at the web portal page. The data object, as a h5ad file, can also be downloaded from the portal page. The UK dataset is available at the European Genome–Phenome Archive under accession number [EGAD00001007718](#). Counts matrices from bronchial brushings obtained from patients at Northwestern Memorial Hospital, Chicago, are available at the Gene Expression Omnibus under accession number [GSE168215](#). As data are from living patients, these data are available under managed data access.

Code availability

All data analysis scripts are available at GitHub (<https://github.com/Teichlab/COVID-19paed>).

References

1. Swann, O. V. et al. Clinical characteristics of children and young people admitted to hospital with covid-19 in United Kingdom: prospective multicentre observational cohort study. *Brit. Med. J.* **370**, m3249 (2020).
2. Castagnoli, R. et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in children and adolescents: a systematic review. *JAMA Pediatr.* **174**, 882–889 (2020).
3. Ledford, H. Deaths from COVID ‘incredibly rare’ among children. *Nature* **595**, 639–639 (2021).
4. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
5. Pang, L. et al. Influence of aging on deterioration of patients with COVID-19. *Aging* **12**, 26248–26262 (2020).

6. Muus, C. et al. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* **27**, 546–559 (2021).
7. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
8. Bunyavanich, S., Do, A. & Vicencio, A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. *JAMA* **323**, 2427–2429 (2020).
9. Saheb Sharif-Askari, N. et al. Airways expression of SARS-CoV-2 receptor, ACE2, and TMPRSS2 is lower in children than adults and increases with smoking and COPD. *Mol. Ther. Methods Clin. Dev.* **18**, 1–6 (2020).
10. Koch, C. M. et al. Age-related differences in the nasal mucosal immune response to SARS-CoV-2. *Am. J. Respir. Cell Mol. Biol.* <https://doi.org/10.1165/rcmb.2021-0292OC> (2021).
11. Loske, J. et al. Pre-activated antiviral innate immunity in the upper airways controls early SARS-CoV-2 infection in children. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01037-9> (2021).
12. Schultze, J. L. & Aschenbrenner, A. C. COVID-19 and the human innate immune system. *Cell* **184**, 1671–1692 (2021).
13. Schoggins, J. W. Interferon-stimulated genes: what do they all do? *Annu. Rev. Virol.* **6**, 567–584 (2019).
14. Ziegler, C. G. K. et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *Cell* **184**, 4713–4733 (2021).
15. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
16. Wang, E. Y. et al. Diverse functional autoantibodies in patients with COVID-19. *Nature* **595**, 283–288 (2021).

17. Major, J. et al. Type I and III interferons disrupt lung epithelial repair during recovery from viral infection. *Science* **369**, 712–717 (2020).
18. Broggi, A. et al. Type III interferons disrupt the lung epithelial barrier upon viral recognition. *Science* **369**, 706–712 (2020).
19. Berlin, D. A., Gulick, R. M. & Martinez, F. J. Severe COVID-19. *N. Engl. J. Med.* **383**, 2451–2460 (2020).
20. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
21. Wilk, A. J. et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat. Med.* **26**, 1070–1076 (2020).
22. Zhang, J.-Y. et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* **21**, 1107–1118 (2020).
23. Stephenson, E. et al. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **27**, 904–916 (2021).
24. Pierce, C. A. et al. Immune responses to SARS-CoV-2 infection in hospitalized pediatric and adult patients. *Sci. Transl. Med.* **12**, eabd5487 (2020).
25. Weisberg, S. P. et al. Distinct antibody responses to SARS-CoV-2 in children and adults across the COVID-19 clinical spectrum. *Nat. Immunol.* **22**, 25–31 (2021).
26. Deprez, M. et al. A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* **202**, 1636–1645 (2020).
27. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).
28. Chua, R. L. et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).

29. Grant, R. A. et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641 (2021).
30. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016–1035 (2020).
31. Wang, W. et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* **323**, 1843–1844 (2020).
32. Yu, F. et al. Quantitative detection and viral load analysis of sars-cov-2 in infected patients. *Clin. Infect. Dis.* **71**, 793–798 (2020).
33. Vieira Braga, F. A. et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 (2019).
34. Zhu, N. et al. Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nat. Commun.* **11**, 3910 (2020).
35. Fang, Y. et al. Distinct stem/progenitor cells proliferate to regenerate the trachea, intrapulmonary airways and alveoli in COVID-19 patients. *Cell Res.* **30**, 705–707 (2020).
36. Ruiz García, S. et al. Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, dev177428 (2019).
37. Ygberg, S. & Nilsson, A. The developing immune system—from foetus to toddler. *Acta Paediatr.* **101**, 120–127 (2012).
38. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045 (2020).
39. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest.* **130**, 2620–2629 (2020).

40. Aschenbrenner, A. C. et al. Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med.* **13**, 7 (2021).
41. Silvin, A. et al. Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* **182**, 1401–1418 (2020).
42. Galani, I.-E. et al. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat. Immunol.* **22**, 32–40 (2021).
43. Lee, J. S. et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci. Immunol.* **5**, eabd1554 (2020).
44. Diebold, S. S. et al. Viral infection switches non-plasmacytoid dendritic cells into high interferon producers. *Nature* **424**, 324–328 (2003).
45. Saichi, M. et al. Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. *Nat. Cell Biol.* **23**, 538–551 (2021).
46. Zhou, R. et al. Acute SARS-CoV-2 infection impairs dendritic cell and T cell responses. *Immunity* **53**, 864–877 (2020).
47. Lokugamage, K. G. et al. Type I interferon susceptibility distinguishes SARS-CoV-2 from SARS-CoV. *J. Virol.* **94**, e01410-20 (2020).
48. Schurink, B. et al. Viral presence and immunopathology in patients with lethal COVID-19: a prospective autopsy cohort study. *Lancet Microbe* **1**, e290–e299 (2020).
49. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T cell development, localization, and function throughout life. *Immunity* **48**, 202–213 (2018).

50. Naylor, K. et al. The influence of age on T cell generation and TCR diversity. *J. Immunol.* **174**, 7446–7452 (2005).
51. Hagai, T. et al. Gene expression variability across cells and species shapes innate immunity. *Nature* **563**, 197–202 (2018).
52. Li, S. et al. SARS-CoV-2 triggers inflammatory responses and cell death through caspase-8 activation. *Signal Transduct. Target. Ther.* **5**, 235 (2020).
53. Lee, P. Y. et al. Distinct clinical and immunological features of SARS-CoV-2-induced multisystem inflammatory syndrome in children. *J. Clin. Invest.* **130**, 5942–5950 (2020).
54. Worlock, K. B. Cell dissociation from nasal, bronchial and tracheal brushings with cold-active protease for single-cell RNA-seq. *protocols.io* <https://doi.org/10.17504/protocols.io.btpunmnw> (2021).
55. Lun, A. T. L. et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).
56. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
57. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
58. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
59. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
60. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

61. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**,giaa151 (2020).
62. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
63. Wood, D. kraken2, <https://github.com/DerrickWood/kraken2> (2018).
64. Bost, P. Viral-Track, <https://github.com/PierreBSC/Viral-Track> (2020).
65. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
67. Quinlan, A. bedtools2, <https://github.com/arq5x/bedtools2> (2021).
68. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
69. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
70. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
71. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
72. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).

73. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
74. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
75. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2020).
76. Reimand, J. et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
77. Cantuti-Castelvetri, L. et al. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856–860 (2020).
78. Daly, J. L. et al. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861–865 (2020).
79. Wang, K. et al. CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. *Signal Transduct. Target. Ther.* **5**, 283 (2020).
80. Tang, X. et al. Transferrin receptor is another receptor for SARS-CoV-2 entry. Preprint at <https://doi.org/10.1101/2020.10.23.350348> (2020).
81. Young, A. M. H. et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat. Genet.* **53**, 861–868 (2021).
82. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).

Acknowledgements

We acknowledge assistance from L. Thorne, P. S. Chia, R. Hynds, J. Eliasova, D. King, M. Heightmann, M. Marks, M. Avari, T. Mistry, M. Shaw-Taylor, R. Pereira, J. Machta, J. Lim, R. Prendecki, C. Frauenfelder, J. Rudd, A. Hall and the staff at the Sanger Institute Core Sequencing facility. We thank R. Jenner and the staff at the UCLH/UCL Biomedical

Research Centre for the use of their 10x Chromium controller. We acknowledge funding from Wellcome (WT211276/Z/18/Z and Sanger core grant WT206194). M.Z.N., S.M.J. and K.B.M. have been funded by the Rosetrees Trust (M944, M35-F2) and from Action Medical Research (GN2911). This project has been made possible in part by grants 2017-174169 and 2019-202654 from the Chan Zuckerberg Foundation and has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 874656. M.Z.N. acknowledges funding from the Rutherford Fund Fellowship allocated by the MRC, and M.Z.N. and S.M.J. from the UK Regenerative Medicine Platform 2 (MR/5005579/1), the Longfonds BREATH consortium and University College London Hospitals Biomedical Research Centre. M.Y. is funded by The Jikei University School of Medicine. K.B.W. acknowledges funding from University College London, Birkbeck MRC Doctoral Training Programme. C.M.S. and M.Z.N. acknowledge support from BBSRC (BB/V006738/1). We acknowledge support from the NIHR Great Ormond Street Biomedical Research Centre and the Great Ormond Street Hospital Children's Charity. S.S. was supported by a Japan Society for the Promotion of Science Overseas Fellowship (310072). R.G.W. was supported by NIH grant U19AI135964 and a GlaxoSmithKline Distinguished Scholar in Respiratory Health grant from the CHEST Foundation. A.V.M. was supported by NIH grant U19AI135964. This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications/).

Author information

Author notes

1. These authors contributed equally: Masahiro Yoshida, Kaylee B. Worlock, Ni Huang, Rik G. H. Lindeboom
2. These authors jointly supervised this work: Marko Z. Nikolić, Kerstin B. Meyer

Affiliations

1. UCL Respiratory, Division of Medicine, University College London, London, UK

Masahiro Yoshida, Kaylee B. Worlock, Josephine L. Barnes, Jessica Allen-Hyttinen, Henry Yung, Puja Mehta, Iván T. Herczeg, Sam M. Janes & Marko Z. Nikolić

2. Division of Respiratory Diseases, Department of Internal Medicine, Jikei University School of Medicine, Tokyo, Japan

Masahiro Yoshida

3. Wellcome Sanger Institute, Cambridge, UK

Ni Huang, Rik G. H. Lindeboom, Natsuhiko Kumasaka, Cecilia Dominguez Conde, Lira Mamanova, Liam Bolt, Laura Richardson, Krzysztof Polanski, Elo Madissoon, Anna Wilbrey-Clark, Waradon Sungnak, J. Patrick Pett, Juliane Weller, Elena Prigmore, Muzlifah Haniffa, Menna R. Clatworthy, Sarah A. Teichmann & Kerstin B. Meyer

4. NIHR Great Ormond Street BRC and UCL Institute of Child Health, London, UK

Colin R. Butler, Brendan C. Jones, Soichi Shibuya, Christopher O'Callaghan, Neil J. Sebire, Paolo De Coppi & Claire M. Smith

5. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Colin R. Butler, Brendan C. Jones, Christopher O'Callaghan, Neil J. Sebire & Paolo De Coppi

6. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK

Elo Madissoon

7. University College London Hospitals NHS Foundation Trust, London, UK

Eliz Kilich, Angus de Wilton, Henry Yung, Puja Mehta, Jonathan M. Cohen, Sam M. Janes & Marko Z. Nikolić

8. Royal Free Hospital NHS Foundation Trust, London, UK

Aarash Saleh, Anita Saigal, Vivian Chu, Clare Cane & Aikaterini Iordanidou

9. UCL Division of Infection and Immunity, University College London, London, UK

Ann-Kathrin Reuschl, G. R. Scott Budinger & Clare Jolly

10. Division of Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

A. Christine Argento, Richard G. Wunderink, Sean B. Smith, Taylor A. Poor, Catherine A. Gao, Jane E. Dematte, Helen K. Donnelly, Nikolay S. Markov, Ziyan Lu & Alexander V. Misharin

11. Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK

Gary Reynolds & Muzlifah Haniffa

12. Department of Medicine, University of Cambridge, Cambridge, UK

Georgina S. Bowyer, Matthew Coates & Menna R. Clatworthy

13. Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Matthew Coates

14. Wellcome, MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

Fernando J. Calero-Nieto & Berthold Göttgens

15. Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge, UK

Sarah A. Teichmann

Consortia

NU SCRIPT Study Investigators

- A. Christine Argento
- , Catherine A. Gao
- , Alexander V. Misharin
- , G. R. Scott Budinger
- , Jane E. Dematte
- , Helen K. Donnelly
- , Nikolay S. Markov
- , Richard G. Wunderink
- , Sean B. Smith
- , Taylor A. Poor
- & Ziyan Lu

Contributions

M.Z.N. and K.B.M. conceived, set up and directed the study. C.R.B., E.K., A.d.W., B.C.J., A. Saleh, A. Saigal, H.Y., S.M.J., S.S., P.M., N.J.S., P.d.C., V.C., J.M.C., C.C., A.I. and M.Z.N. recruited patients, and collected samples (where applicable also through bronchoscopies) and clinical metadata. K.B.W. and M.Y. assisted with sample and metadata collection, isolated PBMCs and performed single-cell isolation of nasal, tracheal and bronchial brushings. K.B.W. and M.Y. performed 10x and CITE-seq and isolated DNA for genotyping. J.A.-H. collected samples, performed single-cell isolation and 10x (including CITE-seq) on post-COVID-19 samples. J.L.B. and I.T.H. helped with the study set-up, CITE-seq and isolated DNA for genotyping. L.M., L.B. and L.R. prepared sequencing libraries and conducted the sequencing. E.P. coordinated sample shipment and metadata

collection. N.H., R.G.H.L., N.K., C.D.C., E.M., K.P., J.P.P. and J.W. performed bioinformatics analysis. M.Z.N., K.B.M., K.B.W., M.Y., R.G.H.L., N.H., N.K., C.D.C., E.M. and W.S. interpreted the data. K.P. facilitated online data hosting. G.R. and M.H. provided help with PBMC annotation. N.J.S., B.C.J. and S.S. provided stored healthy paediatric control nasal tissue blocks. M.C., A.W.-C., G.S.B. and M.R.C. performed experiments to collect and stain biopsies. F.J.C.-N. and B.G. designed the CITE-seq panel and advised on CITE-seq experimental design. K.B.M., M.Z.N., K.B.W., M.Y., R.G.H.L. and N.H. wrote the manuscript. E.M., S.A.T., B.G., W.S., S.M.J., J.P.P. and L.M. edited the manuscript. C.M.S., C.O., P.d.C., S.M.J. and C.R.B. provided support through ethics and patient recruitment. C.J. and A.-K.R. provided support in setting up and training for all CL3 work. For the Northwestern University (bronchial) samples, A.C.A., C.A.G., G.R.S.B., J.E.D., R.G.W., S.B.S. and T.A.P. performed bronoscopies, collection of bronchial brushings and curation of clinical metadata. A.V.M. performed sample processing and analysis. H.K.D. obtained informed consent and coordinated sample collection. N.S.M. performed analysis. Z.L. performed sample processing and library construction.

Corresponding authors

Correspondence to [Marko Z. Nikolić](#) or [Kerstin B. Meyer](#).

Ethics declarations

Competing interests

In the past three years, S.A.T. has worked as a consultant for Genentech, Roche and Transition Bio, and is a remunerated member of the Scientific Advisory Boards of Qiagen, GlaxoSmithKline and Foresite Labs and an equity holder of Transition Bio. P.M. is a Medical Research Council-GlaxoSmithKline (MRC-GSK) Experimental Medicine Initiative to Explore New Therapies (EMINENT) clinical training fellow with project funding, has served on an advisory board for SOBI, outside the submitted work, and receives co-funding by the National Institute for Health Research (NIHR)

University College London Hospitals Biomedical Research Centre (UCLH BRC).

Peer review

Peer review information

Nature thanks Ross Levine and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Overview of patient cohort.](#)

(a) Overview of samples taken in our healthy, COVID-19 and post-COVID-19 cohorts. COVID-19 severity was classified as asymptomatic, mild (symptomatic without oxygen requirement or respiratory support), moderate (requiring oxygen without respiratory support) or severe (requiring non-invasive or invasive ventilation). Post-COVID-19 patients were sampled 3 months after recovering from severe COVID-19. (b) Timeline of sample collections from COVID-19 positive (18 adults and 19 paediatric) and post-COVID-19 (13 adults and 2 paediatric) patients enrolled in our study. Sample collections are shown relative to symptom onset and a SARS-CoV-2 positive RT-qPCR test, to which all patients are aligned.

[Extended Data Fig. 2 Airway single-cell metadata, proportions and cell type markers.](#)

(a) UMAP visualization of annotated airway scRNA-seq dataset from Fig. 1b coloured by COVID-19 status and age groups. **(b)** Bar plot comparing nasal epithelial cell type compositions across COVID-19 status and age groups. **(c)** Dot plots showing marker genes for annotated airway epithelial and immune cell types, with fraction of expressing cells and average expression within each cell type indicated by dot size and colour, respectively.

Extended Data Fig. 3 Supplementary information for airway cell type annotation.

(a) Detailed marker genes for distinct airway myeloid populations in our data set listing marker genes that are unique to each of the defined populations, whilst markers that are common to closely related myeloid cell types are shown on the right side of the panel. **(b)** Comparison of annotated cell types to published data sets. Marker genes for the three populations identified as differentiating to ciliated cells²⁸ and markers of transit epithelial cells (Transit epi 1 and 2). Deu; deuterosomal, Ba-d; basal differentiating, IRC; interferon responsive cell. **(c, d)** Logistic regression based label transfer for the data sets in (c) Chua et al²⁸ and (d) Ziegler et al¹⁴. **(e)** Bar chart showing changes in nasal epithelial cell type proportions observed across age within our paediatric and adult healthy cohorts. Error bars indicate two times standard error of the mean.

Extended Data Fig. 4 Expression of viral entry-associated genes in the airways.

(a) Dot plots showing cell type expression of viral entry-associated genes within the upper airways of healthy adults ($n = 7$), healthy children ($n = 30$), COVID-19 adults ($n = 10$) and COVID-19 children ($n = 18$) respectively, included genes linked to SARS-CoV-2, SARS-CoV, MERS-CoV, Rhinovirus-C and Influenza A infections. The fraction of expressing cells and average expression within each cell type is indicated by dot size and colour, respectively. **(b)** Spearman correlation between the fraction of cells with detected viral RNA and the average expression of entry factors, as in (a), across cell types within the airways of COVID-19 patients samples

(with viral reads ≥ 5) within 5 days of a positive SARS-CoV-2 qPCR test (Early) and those sampled longer than 5 days prior to onset of symptoms or positive SARS-CoV-2 qPCR test, whichever was longer (Late). Dots in blue indicate $p < 0.05$. (c) Expression of ACE2 in paediatric airway cells in each cell type averaged by donor (upper) and in each donor (lower) and coloured by COVID-19 status. Error bars indicate two times standard error of the mean across donors. Numbers in brackets indicate numbers of COVID-19 donors/healthy donors.

Extended Data Fig. 5 Airway cell type proportion analysis, interferon responses and differential gene expression.

(a) Dot plot showing fold change and statistical significance of all airway cell type proportions across location of sampling, age group and COVID-19 status, respectively, estimated by fitting Poisson generalized linear mixed models taking into account other technical and biological variables (see [Methods](#)). (b) Feature importance plot depicting the variance accounted for by each of the clinical and technical factors in our statistical analysis of cell type proportions within our airway scRNA-seq dataset. Factors were donor (patient), patients age (Age_bin), sample location (nasal, tracheal, bronchial), COVID-19 status group (COVID-19 positive, negative or post-COVID-19), dataset (UK cohort or Chicago Cohort) sex, 10x chromium 5' single-cell sequencing kit version (kit_version) smoking status (non-smoker, ex-smoke or current), date and other factors (residual). Note: Error bars were not able to be generated for sex, Kit_version and smoker. 97 samples contributed to the estimation of variances and their standard errors. (c) Response to interferon by airway cell type. Scores of GO term gene signatures for the terms: response to type 1 interferon (GO:0035455 or GO:0034340) and interferon-gamma (GO:0034341) across cell types. Scores were calculated with Scanpy as the average expression of the signature genes subtracted with the average expression of randomly selected genes from bins of corresponding expression values. (d) Differential gene expression contrasting COVID-19 and non-COVID-19 samples in transit epithelial 1 cells, inflammatory goblet 2 cells, and mono IL-6 cells.

Extended Data Fig. 6 Expression of cell type markers and immune compartment dynamics.

(a) Expanded dot plot from Fig. 3d showing the RNA expression of cell type marker genes and interferon-stimulated genes. **(b)** Dotplot showing the cell surface protein expression of cell type marker proteins. In both **a** and **b** the size of the dot is scaled to the percentage of cells that have at least one count for each gene or protein, and the colour is scaled to the z-score normalized expression of each gene or protein. **(c)** Comparison of our manual cell type PBMC annotation vs an automated annotation performed by Azimuth. **(d)** Fold changes of immune cell type proportions across age group and disease status. Age and disease specific changes were deconvoluted by fitting Poisson generalized linear mixed models taking into account other confounders such as sex and ethnicity. **(e)** Feature importance plot showing the variance that can be explained by the different features that were included in the Poisson linear mixed model that was fitted on the cell type proportions in the PBMC data. 80 samples contributed to the estimation of variances and their standard errors. **(f)** Bar plots showing the average immune cell proportions in PBMC samples. Cell types are colour coded and grouped based on their age group and disease status. N denotes the amount of samples in each group, while K denotes the amount of cells per group. **(g)** UMAPs as in Fig. 3a in which the COVID-19 status (left panel) and the age group (right panel) is visualized for each cell.

Extended Data Fig. 7 Immune cell population dynamics.

(a) Fractions of unique BCR sequences show the differences in immune repertoire diversity over age and disease. **(b)** UMAP visualization as in Fig. 3a showing the annotated interferon-stimulated subpopulations in clusters 35–42. **(c)** Boxplot showing the percentage of PBMCs that are interferon-stimulated in asymptomatic or symptomatic COVID-19 patients, grouped by the weeks since the onset of symptoms, and separated for adults (left) and children (right). **(d)** Dotplot of Spearman correlations between nasal and blood cell type proportions in paediatric COVID-19 patients and **(e)** in adult COVID-19 patients. In both **d** and **e**, cell type proportions in the nose (x-axis) are compared to the blood (y-axis). Correlations shown in Fig. 3g present a zoom in of the adult panel. Rows and columns in both dotplots are

clustered by hierarchical clustering on the combined matrices. The size of the dots is scaled by the significance of each correlation. Colour is scaled by the Spearman rank-correlation coefficient. If a blood - nose cell type combination shows a positive correlation, this is indicative that if the blood cell type changes in proportion, the nasal cell type changes accordingly, and vice versa. Dots in **a** and **c** represent independent patient samples. Box plots were drawn with the centre line as the median of the data distribution, the hinges as the first and third quartiles, and with the whiskers extending to the lowest and highest values that were within $1.5 \times$ interquartile range of the upper or the lower hinge.

Extended Data Fig. 8 Interferon expression in COVID-19 patient with highest amount of interferon-stimulated blood cells.

(**a**) Ranked barplot and matched dotplots as in Fig. 3h, but showing the expression of all genes that make up the interferon-stimulated gene signature (middle) and the expression of all interferons (right) in all cells, instead of averaged signatures gene expression signatures in specific cell types. (**b**) Dotplot related to Fig. 3h showing the expression of all interferons in all nasal resident (top) and circulating (bottom) cell types that were present in this individual. The size of the dot is scaled to the percentage of cells that have at least one count for each gene or protein, and the colour is scaled to the z-score normalized expression of each gene or protein.

Extended Data Fig. 9 Metagenomic analysis of patient sample reads that were not mapped to the human genome.

(**a**) Dotplot showing the amount of cells that harbour reads aligned to archaea, bacteria, eukaryota (including human reads that initially did not align to the human transcriptome by STARsolo) and viruses. (**b**) Dotplot showing the amount of cells that harbour reads to a selection of disease-relevant bacteria and viruses. Apart from SARS-CoV-2 and non-specific signal found in most samples, we did not detect any pathogens that were highly abundant in samples of interest.

Extended Data Table 1 Summary of patient metadata

Supplementary information

Supplementary Information

Detailed cell type annotation, cell type abbreviations for airway and blood immune cells, IFN production in dendritic cells and additional references.

Reporting Summary

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

Yoshida, M., Worlock, K.B., Huang, N. *et al.* Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature* **602**, 321–327 (2022). <https://doi.org/10.1038/s41586-021-04345-x>

- Received: 06 March 2021
- Accepted: 14 December 2021
- Published: 22 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04345-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/s41586-021-04345-x>

- Article
- Open Access
- [Published: 21 December 2021](#)

Activation mechanism of PINK1

- [Zhong Yan Gan](#) ORCID: [orcid.org/0000-0001-5755-7780](#)^{1,2},
- [Sylvie Callegari](#) ORCID: [orcid.org/0000-0002-4402-945X](#)^{1,2},
- [Simon A. Cobbold](#) ORCID: [orcid.org/0000-0002-9927-8998](#)^{1,2},
- [Thomas R. Cotton](#) ORCID: [orcid.org/0000-0001-6709-9218](#)^{1,2},
- [Michael J. Mlodzianoski](#) ORCID: [orcid.org/0000-0002-3510-9167](#)^{1,2},
- [Alexander F. Schubert](#) ORCID: [orcid.org/0000-0002-1834-3787](#)³,
- [Niall D. Geoghegan](#) ORCID: [orcid.org/0000-0001-9884-0769](#)^{1,2},
- [Kelly L. Rogers](#) ORCID: [orcid.org/0000-0002-6755-0221](#)^{1,2},
- [Andrew Leis](#)^{1,2},
- [Grant Dewson](#) ORCID: [orcid.org/0000-0003-4251-8898](#)^{1,2},
- [Alisa Glukhova](#) ORCID: [orcid.org/0000-0003-4146-965X](#)^{1,2,4,5} &
- [David Komander](#) ORCID: [orcid.org/0000-0002-8092-4320](#)^{1,2}

Nature volume 602, pages 328–335 (2022)

- 10k Accesses
- 167 Altmetric
- [Metrics details](#)

Subjects

- [Autophagy](#)
- [Cryoelectron microscopy](#)
- [Enzyme mechanisms](#)
- [Kinases](#)
- [X-ray crystallography](#)

Abstract

Mutations in the protein kinase PINK1 lead to defects in mitophagy and cause autosomal recessive early onset Parkinson's disease^{1,2}. PINK1 has many unique features that enable it to phosphorylate ubiquitin and the ubiquitin-like domain of Parkin^{3,4,5,6,7,8,9}. Structural analysis of PINK1 from diverse insect species^{10,11,12} with and without ubiquitin provided snapshots of distinct structural states yet did not explain how PINK1 is activated. Here we elucidate the activation mechanism of PINK1 using crystallography and cryo-electron microscopy (cryo-EM). A crystal structure of unphosphorylated *Pediculus humanus corporis* (*Ph*; human body louse) PINK1 resolves an N-terminal helix, revealing the orientation of unphosphorylated yet active PINK1 on the mitochondria. We further provide a cryo-EM structure of a symmetric *Ph*PINK1 dimer trapped during the process of *trans*-autophosphorylation, as well as a cryo-EM structure of phosphorylated *Ph*PINK1 undergoing a conformational change to an active ubiquitin kinase state. Structures and phosphorylation studies further identify a role for regulatory PINK1 oxidation. Together, our research delineates the complete activation mechanism of PINK1, illuminates how PINK1 interacts with the mitochondrial outer membrane and reveals how PINK1 activity may be modulated by mitochondrial reactive oxygen species.

[Download PDF](#)

Main

Parkinson's disease (PD) is a neurodegenerative disorder in which progressive motor symptoms are caused by the loss of dopaminergic neurons in the substantia nigra pars compacta¹³. There is presently no treatment that slows or stops PD progression. Although PD is typically a disease of people aged 60 and above, one in ten cases of PD occurs early (<50 years) and can commonly be traced to a mutation in one of around 15 PARK genes¹⁴.

Cell biological insights in the past decade have linked many PARK genes to mitochondrial health. *PARK6* (also known as *PINK1*), which encodes the ubiquitin kinase PINK1, and *PARK2* (also known as *PRKN*), which encodes the E3 ubiquitin ligase Parkin, are frequently mutated in autosomal recessive early onset PD (EOPD) and both function in mitophagy—a cell's disposal mechanism for damaged mitochondria^{2,15,16}. The activation of PINK1 is one of the most upstream events in mitophagy. Usually, PINK1 is imported into the mitochondria, cleaved by proteases such as PARL, extracted and degraded by the proteasome^{2,16}. Mitochondrial damage stops PARL cleavage and PINK1 is stabilized on the cytosolic face of the mitochondrial outer membrane (MOM), where it is associated with the translocase of the outer membrane (TOM)^{17,18}. Stabilized PINK1 is activated, leading to

phosphorylation of ubiquitin^{4,5,6,7,8} and recruitment and activation of Parkin^{19,20,21,22}. Active Parkin coats damaged mitochondria with ubiquitin, triggering mitophagy^{15,16}.

Parkin activation has been structurally resolved^{16,21,22}, but it has remained unclear how PINK1 is activated. Human PINK1 (*HsPINK1*) is a divergent Ser/Thr protein kinase that comprises three insertions in the N-lobe as well as a C-terminal extension. The first crystal structures were generated using insect variants that contain only two insertions²³ (Fig. 1a). *Tribolium castaneum* (*Tc*; flour beetle) PINK1 kinase domain structures after extensive engineering^{11,12} adopt a typical bilobal kinase fold with an extended α C helix, but insertions were disordered or had been removed (Extended Data Fig. 1). Our structure of *PhPINK1* bound to a ubiquitin mutant (Ub TVLN) and a nanobody¹⁰ revealed how an autophosphorylation event organizes the N-lobe through an unusual ‘kinked’ α C helix, and orders insertion-3 to form a ubiquitin-binding site¹⁰. All *TcPINK1* and *PhPINK1* structures are in active conformations seemingly poised for catalysis²⁴, yet the differences in conformations and species raised questions of how either related to human PINK1.

Fig. 1: Crystal structure of the cytosolic portion of *PhPINK1*.

 **figure 1**

a, *PhPINK1* constructs that were used previously (Protein Data Bank (PDB): [6EQI](#)¹⁰) and structurally characterized in this study. The mitochondrial-targeting sequence (MTS), outer mitochondrial membrane localization signal (OMS), TMD, N-helix and linker, insertion-2 (i2), insertion-3 (i3) and the CTR are indicated. **b**, The crystal structure of unphosphorylated *PhPINK1*(D334A) (amino acids 115–575), with an extended α C helix and disordered insertion-3. The N-helix (teal) packs against the CTR and directly follows the predicted TMD that interacts with the TOM complex

(not to scale) (Extended Data Figs. 1 and 2 and Supplementary Table 1). aa, amino acids. **c**, EOPD mutations in the N-helix and CTR affect the interface. Mutations according to refs. 2,27. **d**, The previous structure of the *PhPINK1*–Ub TVLN complex (PDB: 6EQI; Ub TVLN is shown in grey, without the nanobody¹⁰), with a kinked α C helix, phosphorylated (p) Ser202 and ordered insertion-3.

Using crystallography and cryo-EM, we resolved the activation mechanism of *PhPINK1* showing (1) the unphosphorylated state; (2) a dimerized state capturing *PhPINK1* just before *trans*-autophosphorylation; (3) a phosphorylated state undergoing a conformational change in the N-lobe; to arrive at (4) the phosphorylated, ubiquitin-binding state of *PhPINK1* (ref. 10). We show that similar structural transitions apply to *HsPINK1* and reveal how PINK1 is not only regulated by phosphorylation, but also by oxidation.

The structure of unphosphorylated *PhPINK1*

To obtain a structure of unphosphorylated *PhPINK1*, we expressed, purified and crystallized the entire cytosolic portion of *PhPINK1* with an inactivating D334A mutation of the catalytic base residue located in the HRD motif (Fig. 1a; kinase nomenclature is shown in Extended Data Fig. 1e). The final 3.53 Å crystal structure of unphosphorylated *PhPINK1* (Fig. 1b, Extended Data Figs. 1 and 2 and Supplementary Table 1) revealed a kinase fold that is most similar to *TcPINK1* structures^{11,12} with an extended α C helix and a disordered insertion-3 (Fig. 1b and Extended Data Fig. 1).

A previously undescribed N-terminal region (amino acids 115–142) extends from the kinase N-lobe and adds a helix (amino acids 121–135, hereafter N-helix) to the kinase C-lobe to extend the C-terminal region (CTR) (Fig. 1b). A structurally reminiscent N-helix–CTR region in pseudokinases SgK223 and SgK269 (Extended Data Fig. 2a) forms a dimerization domain^{25,26}. By contrast, the N-helix in *PhPINK1* generates a monomeric enzyme in solution (Extended Data Fig. 2b–d). The N-helix–CTR interface is a hotspot for EOPD mutations (Fig. 1c) and its structural integrity is crucial for PINK1 function²⁷. Moreover, the N-helix immediately follows the predicted transmembrane domain (TMD, predicted amino acids 101–117; Thr119 is the first visible residue in our structure), suggesting a restrained action radius for PINK1 at the MOM (Fig. 1b).

The structural differences between unphosphorylated *PhPINK1* (Fig. 1b) and the *PhPINK1*–Ub TVLN complex¹⁰ (Fig. 1d) resolve the species disparity, but imply that there are large-scale conformational changes in the N-lobe during PINK1 activation.

PhPINK1 oligomerization enables cryo-EM

A serendipitous observation illuminated the event of *PhPINK1* autophosphorylation. *PhPINK1* with a different kinase-inactivating mutation, D357A in the kinase DFG motif (Extended Data Fig. 1e), eluted in two distinct peaks on size-exclusion chromatography (SEC) during purification (Fig. 2a and Extended Data Fig. 3a). The major peak corresponded to a molecular mass of 604 kDa as measured using SEC multi-angle light scattering (MALS) and suggested the presence of an oligomerized form of *PhPINK1* (molecular mass, 53 kDa). A minor peak of 66 kDa indicated transient dimer formation (Fig. 2a). Furthermore, melting curve analysis of *PhPINK1*(D357A) revealed a high secondary melting temperature that was absent for monomeric *PhPINK1* variants (Extended Data Fig. 3b). Together, these data suggested the formation of a defined *PhPINK1* oligomer.

Fig. 2: Oligomerization of a kinase inactive *PhPINK1* enables cryo-EM.

 **figure 2**

a, SEC–MALS analysis of *PhPINK1* (amino acids 115–575) variants (Fig. [1a](#)). Absorbance was measured at a wavelength of 280 nm. Theoretical and observed molecular mass values are indicated. Each protein displayed identical behaviour in at least three SEC runs, and SEC–MALS experiments were performed twice. **b**, Cryo-EM density map for the *PhPINK1*(D357A) dodecamer at 2.48 Å, indicating monomers

in different colours and dimers in shades of the same colour. Left, top view with three-fold symmetry. Right, side view with two-fold symmetry (Extended Data Fig. 3 and Supplementary Table 2).

Indeed, negative-stain EM revealed that *PhPINK1*(D357A) formed highly regular, easily discernible monodisperse particles (Extended Data Fig. 3c). Subsequent cryo-EM analysis resulted in a 2.48 Å density map of a dodecamer of *PhPINK1*, in which two rings of three dimers form a bagel-like arrangement with D_3 symmetry (Fig. 2b, Methods, Extended Data Figs. 3 and 4 and Supplementary Table 2). A *PhPINK1* dimer is the smallest unit within the oligomer, and symmetry expansion and local refinement of a masked dimer increased the resolution to 2.35 Å (Extended Data Fig. 3d,f).

We speculated that wild-type (WT) *PhPINK1* may not oligomerize due to autophosphorylation¹⁰ and, indeed, when WT *PhPINK1* was dephosphorylated with λ -phosphatase (λ -PP), it formed an oligomer; subsequent rephosphorylation of the oligomer destabilized it (Extended Data Fig. 4c; see below). Dodecamer formation seems to be *PhPINK1* specific, as equivalent constructs of *TcPINK1* did not form oligomers (Extended Data Fig. 4d), and oligomer-enabling interface residues are not conserved in *TcPINK1* and *HsPINK1* (Extended Data Fig. 4e).

The structural basis for autophosphorylation

The *PhPINK1*(D357A) dimer captures the enzyme in the process of *trans*-autophosphorylation, in which a loop of the kinase N-lobe of one monomer is placed into the substrate-binding site of the second monomer in a symmetric contact (Fig. 3a). For this contact to occur, the α C helices are required to be fully extended such that the adjacent Ser202 residue sits in the phosphate-accepting position (Fig. 3a–c). Ser202 forms a hydrogen bond with the catalytic base Asp334 of the HRD motif, typical of kinase-substrate interactions. Modelling ATP into the nucleotide-binding site places the γ -phosphate within 2.3 Å of the Ser202 hydroxyl, poised for phosphoryl transfer¹⁰ (Fig. 3b,c and Extended Data Fig. 5a).

Fig. 3: The *PhPINK1* dimer before *trans*-autophosphorylation.

 **figure 3**

a, The structure of the *PhPINK1*(D357A) dimer in surface representation. Top, front view of the dimer with empty ATP-binding sites. The Ser202-containing loop (Ser202 circled in white) reaches into the acceptor site of the opposing kinase domain. Middle, back view of the dimer, with complementary activation segments (cartoon, coloured). Bottom, side view of the dimer showing the N-helices, indicating how it may sit on the MOM interacting with a TOM complex. A molecular model of the *PhPINK1* dimer manually docked onto the TOM complex is shown in Extended Data Fig. 5d. **b–e**, Detailed views in stick representation. The dotted lines indicate hydrogen bonds. **b**, ATP-binding and *trans*-autophosphorylation interactions. ATP was modelled from PDB [2PHK](#) (ref. ⁴²) as before¹⁰. **c**, Coordination of the Ser202-containing loop. **d**, Dimer interactions through activation segments and αEF–αF loops. **e**, The close proximity of P-loop Cys169 residues during dimer formation (Extended Data Fig. 5e). **f**, Conservation of *PhPINK1* Cys169 in *HsPINK1* (Cys166) but not *TcPINK1* (Thr172). **g**, Formation of a disulphide-linked *HsPINK1* dimer in HeLa cells that were treated with H₂O₂. *HsPINK1* was expressed in HeLa *PINK1*^{−/−} cells and stabilized

with OA treatment ([Methods](#)). H₂O₂ treatment leads to a disulphide-linked *HsPINK1* dimer band visualized on a non-reducing gel, which is absent with C166A mutation, suggesting *HsPINK1* also dimerizes through Cys166. A putative dimer-trapping mutation in *HsPINK1*, D384A (D357A in *PhPINK1*; compare with Fig. [2a](#)), does not further stabilize the dimer. The experiments were performed in biological triplicate with identical results. The uncropped blots are provided in Supplementary Fig. [1](#).

The dimer structure provides profound insights into PINK1 mechanism and regulation. The activation segment (amino acids 357–390) and the adjacent αEF–αF loop (amino acids 393–406) contribute numerous dimer contacts in a symmetric, complementary dimer interface (Fig. [3a](#) (middle) and [3d](#)). Interestingly, the activation segment of PINK1 adopts an identical conformation in all of the structures determined to date (Extended Data Fig. [5b–c](#)).

The dimer structure orients each N-helix–CTR region such that both of the N termini face the MOM (Fig. [3a](#) (bottom)). Strikingly, the distance between TMDs is compatible with placing a TOM complex dimer between the *PhPINK1* N termini (Extended Data Fig. [5d](#)). Although the molecular details of this interaction require refinement, it is tempting to speculate that the TOM complex components^{[17,18,27,28](#)} assist dimerization.

In the *PhPINK1* dimer structure, Cys169 at the turn of the kinase P-loop is located within 3.8 Å of its symmetric counterpart (Fig. [3e](#)). Although Cys at this position is unusual in kinases, it is conserved in *HsPINK1*, but is a Thr in *TcPINK1* (Fig. [3f](#); see below). As a consequence, H₂O₂ treatment and non-reducing gel electrophoresis resolves a disulphide-linked dimer of the oligomer-stabilizing *PhPINK1*(D357A), but not of the WT enzyme, and the C169A mutation prevents oxidative dimerization (Extended Data Fig. [5f](#)). In *TcPINK1*, engineering an equivalent T172C mutation enables oxidative dimerization (Extended Data Fig. [5f](#)), suggesting similar dimer formation in *TcPINK1* (ref. [29](#)).

We next extended our studies to human PINK1 by expressing *HsPINK1* variants in HeLa *PINK1*^{−/−} cells. Treatment with oligomycin/antimycin A (OA)—to stabilize *HsPINK1*—and H₂O₂ leads to a discernible disulphide-linked dimer of WT *HsPINK1* that is prevented by a C166A mutation in *HsPINK1* (equivalent to C169A in *PhPINK1*) (Fig. [3g](#)). Dimerization is seen readily with WT *HsPINK1* without the need of an additional dimer-stabilizing mutation, D384A (equivalent to D357A in *PhPINK1*) (Fig. [3g](#)).

Importantly, multiple EOPD mutations are located in the dimerization interface (Extended Data Fig. [5g](#)). We used oxidative cross-linking on 11 mutants from patients with EOPD, as in Fig. [3g](#), in conjunction with autophosphorylation and

phosphorylated-ubiquitin generation analysis in HeLa *PINK1*^{-/-} cells. Our results show that, despite retaining Cys166, many EOPD mutants are defective in oxidative dimerization, and these mutants display no or distinct autophosphorylation and are deficient in ubiquitin phosphorylation (Extended Data Fig. 5h). We conclude that a similar dimer arrangement is present in *HsPINK1* in which it probably facilitates phosphorylation at Ser228 (Extended Data Fig. 6; further discussed below). Oxidative dimerization of *HsPINK1* could be useful for future studies.

Phosphorylated Ser202 unlocks ubiquitin kinase

The *PhPINK1* dimer structure with Ser202 in the phospho-acceptor site explains the phosphorylation ability and requirements for PINK1 activation. Phos-tag SDS-PAGE kinase assays showed that unphosphorylated, monomeric *PhPINK1*(D334A) was rapidly phosphorylated by WT *PhPINK1* (Extended Data Fig. 6a, b), and phosphorylation was abrogated by a S202A mutation (Extended Data Fig. 6b). Consistently, Ser202 was the only phosphorylation site of *PhPINK1* identified by mass spectrometry³⁰ (Extended Data Fig. 6c). Furthermore, *PhPINK1*(D357A) oligomer is not phosphorylated by WT *PhPINK1* as its acceptor Ser202 is inaccessible in the oligomer (Extended Data Fig. 6d).

To directly test whether phosphorylation of Ser202 is sufficient to confer the phosphorylation activity of *PhPINK1* towards ubiquitin, we purified and dephosphorylated WT *PhPINK1* or *PhPINK1*(S202A). Subsequent co-incubation of *PhPINK1* with both ubiquitin and ATP led to the rapid autophosphorylation of WT *PhPINK1* and ubiquitin phosphorylation, whereas *PhPINK1*(S202A) did not autophosphorylate, nor did it phosphorylate ubiquitin (Extended Data Fig. 6e). These results confirm that, *in vitro*, Ser202 is the main autophosphorylation site of *PhPINK1*, and that autophosphorylation at Ser202 is necessary and sufficient for *PhPINK1* to phosphorylate ubiquitin.

Experiments expressing *HsPINK1* variants in *PINK1*^{-/-} HeLa cells, and assessing both autophosphorylation and ubiquitin phosphorylation, confirmed that phosphorylation of Ser228 (equivalent to *PhPINK1* Ser202) is necessary and sufficient for *HsPINK1* to become a ubiquitin kinase (Extended Data Fig. 6f-i), consistent with previous reports^{30,31,32,33}. We identify that Ser229 and Ser230 are potential additional PINK1 autophosphorylation sites, but phosphorylation at neither site triggers ubiquitin phosphorylation (Extended Data Fig. 6f-i).

Capturing PINK1 conformational changes

We next examined whether phosphorylation of Ser202 leads to a conformation change. To resolve this question and the final step of *PhPINK1* activation from a canonical kinase to a ubiquitin kinase (Fig. 1), we generated a partially cross-linked WT *PhPINK1* oligomer and added Mg²⁺/ATP to phosphorylate each kinase domain (Fig. 4a–c). In the final 3.07 Å cryo-EM map of the dimer (Fig. 4d, Extended Data Fig. 7 and Supplementary Table 2), the kinase C-lobe and the oligomer-forming N-helix–CTR region were highly similar to the original, unphosphorylated structure (Fig. 3). The N-lobe of the kinase showed structural differences, including a break in symmetry leading to different states of the αC helix in the individual kinase domains (Fig. 4d–h).

Fig. 4: The cryo-EM structure of Ser202-phosphorylated *PhPINK1* reveals a conformational change.

 **figure 4**

a, Flowchart for the generation of the phosphorylated and partially cross-linked WT *PhPINK1* dodecamer. RT, room temperature. **b**, Profile of the final SEC run. **c**, Phos-tag analysis of the *PhPINK1* species. The final oligomer fraction comprises homogenously phosphorylated *PhPINK1* and was used for cryo-EM analysis. The experimental workflow shown in **a–c** was performed once in this exact configuration. The uncropped gel is provided in Supplementary Fig. [1](#). **d**, Cryo-EM density map of the Ser202-phosphorylated WT *PhPINK1* dimer at 3.07 Å. A break in symmetry in the

N-lobe is visible ([Methods](#), Extended Data Fig. 7 and Supplementary Table 2). e, EM density for the N-lobe of molecule A shows an extended α C helix (orange) with the phosphorylated Ser202 at the tip. f, The EM density for the N-lobe of molecule B shows a less-ordered state of the α C helix seemingly in transition. g, h, 3D variability analysis enabled the clustering of distinct states of the N-lobe in molecule B. g, In the first cluster, the α C helix (orange) is kinked, and extra density can be modelled by a poly-Ala model of insertion-3 (yellow). h, The second cluster resembles molecule A with an extended α C helix and disordered insertion-3.

In one conformation (molecule A), the α C helix is extended as in the original dimer, with clear density for the phosphorylated Ser202 that remains in the acceptor site of the phosphorylating domain (Fig. 4e and Extended Data Fig. 7e). The ATP-binding site is empty (the nucleotide was probably washed out in the final SEC step), enabling the phosphate to sit in the acceptor site (Extended Data Fig. 7e).

In the second conformation (molecule B), the α C helix and N-lobe appear to be in transition. The α C helix in some particles adopts an extended conformation, while it is kinked in others (Fig. 4f-h). Using 3D variability analysis in cryoSPARC³⁴ we were able to visualize individual conformations and cluster them into two distinct α C conformations (Fig. 4g, h). Importantly, in the conformation with a kinked α C helix, insertion-3 density appears and could be interpreted with a poly-Ala model for this region analogous to the previous PhPINK1–Ub TVLN complex¹⁰ (Fig. 4g).

PINK1 ensures that each kinase domain is phosphorylated^{18,30,33}, explained now by our data (Fig. 4 and Extended Data Fig. 4c). We conceptualized dimer stability as a function of intact Ser202–Asp334 contacts (Extended Data Fig. 8), and investigated different scenarios by assessing oligomer formation. Two intact Ser202–Asp334 contacts enable dimerization and oligomerization of unphosphorylated PhPINK1, whereas unphosphorylated PhPINK1(D334A) mutant or Ser202-phosphorylated PhPINK1 is monomeric, as both contacts are disrupted (Extended Data Fig. 8). Hetero-oligomer formation observed from mixing unphosphorylated PhPINK1(D334A) with Ser202-phosphorylated PhPINK1 showed that one intact Ser202–Asp334 contact is sufficient to enable oligomer formation (Extended Data Fig. 8). We conclude that PINK1 dimers remain stable until each kinase domain is phosphorylated.

Regulation of PINK1 by Cys oxidation

Dimer structures and cross-linking experiments highlighted a possibility for a new mode of PINK1 regulation through oxidation. The cross-linkable Cys169 is one of two conserved and reactive³⁵ Cys residues in the ATP-binding cleft of PINK1 (Fig. 5a and Extended Data Fig. 9a). Biochemical experiments revealed that oxidation with H_2O_2

inhibits ubiquitin kinase activity of phosphorylated, monomeric WT *PhPINK1* (Fig. 5b). *PhPINK1(C169A)* (Cys166 in *HsPINK1*), *PhPINK1(C360A)* (Cys387 in *HsPINK1*) or a double mutant, rendered the kinase less active, but interestingly also rendered it unresponsive to H₂O₂ in vitro (Fig. 5b). *TcPINK1* containing Thr172 instead of Cys169 is less responsive to H₂O₂ inhibition, similar to *PhPINK1(C169A)* (compare Extended Data Fig. 9b with Fig. 5b). Importantly, *PhPINK1* inhibition by H₂O₂ is reversed by dithiothreitol, highlighting the possibility of a regulatory switch (Extended Data Fig. 9c).

Fig. 5: Regulation of PINK1 by oxidation and the model of PINK1 activation.

 figure 5

a, Cys169—which is involved in dimerization (Fig. 3)—and Cys360 line the ATP-binding pocket of *PhPINK1* and are also close to the substrate ubiquitin (PDB: 6EQI)¹⁰. ATP was modelled as in Fig. 3b. **b**, Phos-tag analysis of *PhPINK1*-mediated ubiquitin phosphorylation, with increasing concentrations of H₂O₂. Mutations of Cys169 and Cys360 render *PhPINK1* less active but also unresponsive to H₂O₂. The experiments were performed in biological triplicate with identical results. The uncropped gel is provided in Supplementary Fig. 1. **c**, Conservation of Cys169 and its context in *PhPINK1*, *HsPINK1* and salmon (*Salmo salar*) PINK1 (Extended Data Fig.

9a). d, The workflow for the experiment in **e**. Details are provided in the [Methods](#). Heavy membranes isolated from OA-treated HeLa *PINK1*^{-/-} cells expressing WT *HsPINK1* or *HsPINK1*(C166S/S167N) (SN, mutating the P-loop to a fish-like sequence to generate active Cys166-lacking *HsPINK1*) were pretreated with increasing concentrations of H₂O₂ (as indicated in **e**) to oxidize membrane-associated active *HsPINK1*, before incubation with recombinant ubiquitin and ATP. DTT, dithiothreitol. **e**, Western blotting of the samples generated in **d**, indicating reversible inactivation of *HsPINK1* by H₂O₂, which was not observed using the Cys166-lacking SN mutant. The experiments were performed in biological triplicate. Uncropped blots are provided in Supplementary Fig. [1](#). **f**, Quantification of pSer65-Ub band intensities from experiments in **e** and its repeats ($n = 3$; Supplementary Fig. [2](#)). Intensities were adjusted on the basis of PINK1 levels (reducing gel) and normalized to the 0 mM H₂O₂ condition for each *HsPINK1* variant. Band intensities were quantified using ImageLab (Bio-Rad, v.6.1). **g**, The model for PINK1 activation on the surface of depolarized mitochondria. We expect that α C helix kinking (1) precedes ordering of insertion-3 (2) to form the ubiquitin-binding site. ROS, reactive oxygen species.

[Source data](#)

The relevance of *HsPINK1* Cys166 or Cys387 has not been studied. Consistent with in vitro experiments with *PhPINK1*, *HsPINK1*(C166A) or *HsPINK1*(C387A) had a reduced ability to generate phosphorylated ubiquitin compared with WT *HsPINK1* when expressed in *PINK1*^{-/-} HeLa cells (Extended Data Fig. [9d](#)). Ectopic expression of YFP–Parkin in HeLa cells lacking endogenous Parkin leads to seemingly higher overall levels of phosphorylated ubiquitin that are reduced with *HsPINK1*(C166A) or *HsPINK1*(C387A) (Extended Data Fig. [9e](#)). Consistently, recruitment of YFP–Parkin to the mitochondria is delayed in cells expressing *HsPINK1*(C166A) or *HsPINK1*(C387A) compared with in cells expressing WT *HsPINK1* (Extended Data Fig. [9f,g](#) and Supplementary Video [1](#)).

Although these results revealed the importance of the Cys residues lining the ATP-binding site of *HsPINK1*, the limited ubiquitin kinase activity of *HsPINK1* Cys mutants (Extended Data Fig. [9d–g](#)) precluded studies on the regulation of *HsPINK1* by oxidation. Interestingly, many fish species contain a Ser–Asn motif instead of *HsPINK1* Cys166–Ser167 (Fig. [5c](#) and Extended Data Fig. [9a](#)), and indeed S167N mutation rescues the activity of the impaired *HsPINK1*(C166S) mutant (Extended Data Fig. [9h](#)). The active *HsPINK1* variant lacking Cys166, *HsPINK1*(C166S/S167N) (SN), enabled us to test whether *HsPINK1* can also be reversibly oxidized by enriching for mitochondria from OA-treated HeLa cells expressing *HsPINK1* variants, subjecting them to oxidation and assessing the phosphorylation of recombinant ubiquitin (Fig. [5d–f](#)). WT *HsPINK1* was reversibly inactivated by H₂O₂ treatment, whereas the *HsPINK1* SN mutant was unresponsive to H₂O₂, consistent with

experiments with recombinant *PhPINK1*. Collectively, our data reveal that Cys166 is a site of reversible oxidation in *HsPINK1*, which would enable reactive oxygen species to regulate PINK1 activity (Fig. 5g).

Implications for human PINK1

This completes our model of PINK1 activation and regulation (Fig. 5g). PINK1 stabilized on mitochondria is an active protein kinase even in its unphosphorylated state. Dimerization, possibly with the assistance of the TOM complex, enables *trans*-autophosphorylation at *PhPINK1* Ser202 (Ser228 in *HsPINK1*). Our structural work strongly supports the notion that phosphorylation triggers conformational changes in the N-lobe, including kinking of the α C helix and organization of insertion-3. The dimer is destabilized after both copies of PINK1 are phosphorylated, enabling phosphorylated PINK1 to act as a monomeric ubiquitin kinase and Parkin Ubl kinase to initiate mitophagy (Fig. 5g). We further identified an oxidation switch in Cys169 (Cys166 in *HsPINK1*) that prevents PINK1 activation by potentially preventing dissociation of the autophosphorylated PINK1 dimer and/or hindering ubiquitin access to PINK1 in the phosphorylated, monomeric conformation. A conceptually similar redox switch exists in fructosamine-3-kinases³⁶ (Extended Data Fig. 9i). Considering that PINK1 is located at one of the hotspots of reactive oxygen species production in the cell, this new potential regulatory mechanism warrants further investigation.

Structure predictions from AlphaFold2 (refs. 37,38) for *HsPINK1* (Extended Data Fig. 10) provide further support for our model of PINK1 activation. AlphaFold2 predicts the kinase in a ubiquitin-binding-competent conformation with a kinked α C helix and ordered insertion-3, suggesting that there is ubiquitin kinase activity even in the absence of PINK1 phosphorylation (Extended Data Fig. 10a). This is questionable as Ser228 phosphorylation is a prerequisite for PINK1 ubiquitin kinase activity (Extended Data Fig. 6). We used ColabFold³⁹ to predict a dimeric *HsPINK1* structure (Extended Data Fig. 10c). The result was almost indistinguishable from the dimer arrangement that we experimentally determined in *PhPINK1* (Extended Data Fig. 10d), and our oxidative cross-linking experiments are consistent with the predicted dimer (Fig. 3g). As *HsPINK1* autophosphorylates at Ser228 (Extended Data Fig. 6) and the dimer forms analogously to *PhPINK1*, we anticipate that unphosphorylated *HsPINK1* has an extended α C helix that places Ser228 into the active site of the dimeric molecule to facilitate autophosphorylation. The predicted AlphaFold2 model displays a kinked α C helix and phosphorylated-Ser228-induced organization of insertion-3 and is probably the final phosphorylated state of the ubiquitin kinase PINK1.

Thus, we conclude that the activation mechanism of PINK1 derived here applies across species. Important additional implications include (1) our demonstration that

PINK1 is an active kinase without phosphorylation, reopening the investigation for PINK1 substrates and roles before autophosphorylation⁴⁰; (2) the orientational restraints that PINK1 experiences at the MOM, which would limit its activity radius to MOM-proximal substrates; and (3) the observed regulation by oxidation, which warrants further studies. Our mechanistic insights will probably help efforts to use PINK1 as a drug target to stimulate mitophagy and treat PD⁴¹.

Methods

Cloning

*Ph*PINK1 and *Tc*PINK1 constructs for bacterial expression were cloned into the pOPINK vector⁴³ using In-Fusion Cloning (Takara Bio) incorporating an N-terminal GST tag and 3C cleavage site. Mutagenesis was performed using the Q5 site-directed mutagenesis kit (NEB).

Protein expression and purification

GST-tagged *Ph*PINK1 and *Tc*PINK1 constructs were transfected into *Escherichia coli* Rosetta2 (DE3) pLacI cells (Novagen) and cells were grown at 37 °C in 2× YT medium until an optical density at 600 nm (OD_{600}) of 0.6–0.8 was reached. Protein expression was induced by the addition of 200 μM IPTG and cultures were incubated overnight at 18 °C. Cells were collected by centrifugation at 5,000g for 15 min at 4 °C and frozen at –80 °C. Cells were thawed and lysed by sonication in purification buffer (25 mM Tris (pH 8.5), 300 mM NaCl, 5% (v/v) glycerol, 10 mM DTT) supplemented with EDTA-free protease inhibitor cocktail tablets (Roche), lysozyme and DNase I. Lysates were clarified by centrifugation at 44,000g for 30 min at 4 °C and the supernatant was incubated with Glutathione Sepharose 4B resin (Cytiva). After washing with purification buffer, the resin was incubated with GST–3C PreScission Protease overnight to cleave the GST tag. The cleaved protein was concentrated and purified by SEC using a HiLoad 16/600 Superdex 200 pg column (Cytiva) in SEC buffer (25 mM Tris (pH 8.5), 150 mM NaCl, 5% (v/v) glycerol, 10 mM DTT). The fractions containing pure protein were pooled, concentrated, flash-frozen in liquid nitrogen and stored at –80 °C. *Ph*PINK1 (amino acids 143–575) and GST–*Ph*PINK1 (amino acids 115–575) were purified as described previously^{8,10,21}. For *Tc*PINK1 mutants, additional NaCl was added after SEC to a final concentration of 300 mM NaCl to minimize precipitation during concentration.

His-tagged λ-PP was expressed as *Ph*PINK1 but in BL21(DE3) cells. Cells were lysed by sonication in binding buffer (50 mM Tris (pH 7.4), 500 mM NaCl, 10% (v/v) glycerol) supplemented with EDTA-free protease inhibitor cocktail tablets (Roche),

lysozyme, DNase I, and 1 mM PMSF. Lysates were clarified by centrifugation at 44,000g for 30 min at 4 °C and the supernatant was incubated with HisPur Cobalt Resin (Thermo Fisher Scientific). Resin was washed extensively with binding buffer and subsequently eluted with elution buffer (50 mM Tris (pH 7.4), 10% (v/v) glycerol, 250 mM imidazole). Eluted protein was then purified by SEC using a HiLoad 16/600 Superdex 75 pg column (Cytiva) in 50 mM Tris (pH 7.4), 10% (v/v) glycerol followed by anion exchange chromatography using a Resource Q 6 ml column (Cytiva). The fractions containing pure protein were pooled and concentrated. Glycerol was added to the protein to a final concentration of 50% (v/v) glycerol before flash-freezing in liquid nitrogen.

Crystallization of *PhPINK1(D334A)*

Crystallization screens were performed at the CSIRO C3 Crystallisation Centre. *PhPINK1(D334A)* (amino acids 115–575) was crystallized at 1.5 mg ml⁻¹ by sitting-drop vapour diffusion against 20% (w/v) polyethylene glycol 3350, 0.2 M triammonium citrate, 0.1 M ammonium sulphate, 0.01 M magnesium chloride from 1:1 protein to mother liquor ratio, in 300 nl drops at 8 °C. Needle-like crystals appeared within a week and grew over the following two weeks. Crystals were cryo-protected in mother liquor diluted with 100% (v/v) ethylene glycol to achieve a final concentration of 25% (v/v) ethylene glycol before vitrification in liquid nitrogen.

Crystallographic data collection, phasing and refinement

Diffraction data were collected at the Australian Synchrotron (ANSTO) using the MX2 beamline ($\lambda = 0.9537 \text{ \AA}$, 100 K)⁴⁴ and processed using XDSMe (v.0.6.5.2)⁴⁵. Data were merged using AIMLESS (v.0.5.21)⁴⁶ implemented in CCP4i (v.7.0.001)⁴⁷ and molecular replacement was performed using PHASER (v.2.8.3)⁴⁸ implemented in Phenix (v.1.19.2-4158)⁴⁹ using *PhPINK1* from the *PhPINK1-Ub* TVLN complex (PDB: [6EQI](#))¹⁰ as the search model. Model building was performed in Coot (v.0.9)⁵⁰ and underwent multiple rounds of refinement in Phenix. The N-helix was built de novo. The final model has excellent geometry, with final Ramachandran statistics: 95.09% favoured, 4.65% allowed and 0.26% outliers. Several regions within *PhPINK1* could not be modelled due to disorder; these included amino acids 115–118 (N terminus), 138–146 (N-helix linker to the N-lobe), 182–188 (insertion-2), 260–283 (insertion-3), 493–494 and 526–539 (loops in CTR), and 574–575 (C terminus). Data collection and refinement statistics are provided in Supplementary Table 1.

SEC–MALS

Size-exclusion chromatography multi-angle light scattering (SEC–MALS) experiments were performed using a Superdex 200 Increase 10/300 GL column (Cytiva) coupled with DAWN HELEOS II light scattering detector and Optilab T-rEX refractive index detector (Wyatt Technology). The system was equilibrated in 25 mM Tris (pH 8.5), 150 mM NaCl, 5% (v/v) glycerol, 2 mM TCEP and calibrated using bovine serum albumin (2 mg ml⁻¹) before analysis of experimental samples. For each experiment, 100 µl of purified protein (1 mg ml⁻¹) was injected onto the column and eluted at a flow rate of 0.5 ml min⁻¹. Experimental data were collected and processed using ASTRA (Wyatt Technology, v.7.3.19).

Thermal denaturation assay

Assays were performed with 4 µM *PhPINK1* and 5× SYPRO Orange (Invitrogen) in 25 mM Tris (pH 8.5), 150 mM NaCl, 10 mM DTT at a total volume of 25 µl. Melting curves were recorded in duplicate using a Rotor-Gene Q (Qiagen) with a temperature ramp of 1 °C min⁻¹ from 25 °C to 80 °C. Data were collected using the Rotor-Gene Q Series Software (v.2.3.1) and analysed using GraphPad Prism (v.9.0.0). Melting curves were fitted with Boltzmann sigmoidal functions to calculate the melting temperatures.

Assessment of WT *PhPINK1* oligomerization and preparation of monomeric phosphorylated *PhPINK1*

WT *PhPINK1* (amino acids 115–575, autophosphorylated) at 15 mM was dephosphorylated overnight at 4 °C with 7.5 µM λ-PP in SEC buffer supplemented with 2 mM MnCl₂, and oligomerization was assessed by SEC using a HiLoad 16/600 Superdex 200 pg column in SEC buffer. To assess the effect of *PhPINK1* autophosphorylation on oligomer formation, 10 mM MgCl₂ and 10 mM ATP was added to dephosphorylated *PhPINK1* for 1 min at room temperature to initiate autophosphorylation. The kinase reaction was quenched with 20 mM EDTA, and the oligomeric state of phosphorylated *PhPINK1* was assessed by SEC using a HiLoad 16/600 Superdex 200 pg column in SEC buffer. To prepare monomeric phosphorylated *PhPINK1*, the fractions corresponding to monomeric protein from the SEC run were pooled, concentrated, flash-frozen in liquid nitrogen and stored at –80 °C.

Preparation of *PhPINK1* and phosphorylated *PhPINK1* oligomer for EM analysis

PhPINK1(D357A) (amino acids 115–575) was purified as described above but with the following modifications. An additional anion exchange step using a Mono Q 5/50 GL column (Cytiva) was included before a final SEC run and, during SEC, the protein was buffer exchanged into either SEC buffer (for negative stain EM) or glycerol-free

SEC buffer (for cryo-EM). In both cases, *PhPINK1*(D357A) elutes largely as an oligomer close to but not overlapping with the void volume of a HiLoad 16/600 Superdex 200 pg column. SDS-PAGE analysis of individual fractions was used and oligomer-containing fractions were pooled if more than 95% pure.

To generate the phosphorylated *PhPINK1* oligomer, WT *PhPINK1* (amino acids 115–575) was first dephosphorylated as described above. Dephosphorylated *PhPINK1* was purified by SEC using a HiLoad 16/600 Superdex 200 pg column in DTT-free SEC buffer. The fractions containing oligomeric *PhPINK1* were pooled and immediately treated with 2 mM H₂O₂ for 15 min at room temperature to cross-link the dimer and stabilize the oligomer, and the reaction was quenched with 10 U ml⁻¹ catalase (Sigma-Aldrich). Homogeneous phosphorylation at Ser202 was achieved by the addition of 10 mM MgCl₂ and 10 mM ATP for 1 min at room temperature followed by the addition of 20 mM EDTA to inactivate the kinase and residual λ-PP. Oligomeric phosphorylated *PhPINK1* was purified in a final SEC run on a HiLoad 16/600 Superdex 200 pg column in glycerol- and DTT-free SEC buffer (25 mM Tris (pH 8.5), 150 mM NaCl). The fractions containing the oligomer were pooled, concentrated, flash-frozen in liquid nitrogen and stored at –80 °C.

Negative-stain EM

Negative stain EM data collection was performed at the Ian Holmes Imaging Centre at the Bio21 Molecular Science and Biotechnology Institute, University of Melbourne. Samples were diluted to 0.005–0.01 mg ml⁻¹ and applied to a glow-discharged carbon-coated Cu grid (200 mesh). After 60 s, the solution was blotted off and the grid was stained in 0.8% (w/v) uranyl formate solution for 30 s. Excess solution was blotted off and was followed by two washes in water. The grids were imaged at room temperature using the Talos L120C electron microscope at a magnification of ×52,000 and a defocus value of around –1 μm with a pixel size of 2.44 Å. Particle picking, extraction and 2D classification were performed using RELION (v.3.1)⁵¹.

Cryo-EM sample preparation and data acquisition

Cryo-EM data collection was performed at the Ian Holmes Imaging Centre at the Bio21 Molecular Science and Biotechnology Institute, University of Melbourne. Grids were prepared by dispensing 4 μl of *PhPINK1*(D357A) (2 mg ml⁻¹) or phosphorylated *PhPINK1* (1.3 mg ml⁻¹) onto a glow-discharged UltrAuFoil R1.2/1.3 (300 mesh) or Quantifoil R1.2/1.3 grid (200 mesh), respectively, at 4 °C and 100% humidity. Grids were blotted for 4 s with a nominal blot force of –1 before plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). Data for *PhPINK1*(D357A) were collected on a Thermo Scientific Titan Krios G4 microscope

equipped with a Gatan K3 detector and Biocontinuum energy filter at 300 keV. The acquisition was performed in EFTEM NanoProbe mode, nominal magnification $\times 105,000$, zero-loss slit 10 eV in correlated double-sampling mode (formerly ‘super-resolution mode’, $0.4165 \text{ \AA px}^{-1}$), with a total exposure of 50 e \AA^{-2} over 40 frames. A total of 644 image stacks was collected. The data for phosphorylated *PhPINK1* were collected on the Talos Arctica microscope equipped with a Gatan K2 detector and Bioquantum energy filter at 200 keV. The acquisition was performed in EFTEM NanoProbe mode, nominal magnification $\times 165,000$, zero-loss slit 10 eV and electron-counting mode (0.78 \AA px^{-1}), with a total exposure of 50 e \AA^{-2} over 40 frames. A total of 1,717 image stacks was collected. All data were collected using EPU software (Thermo Fisher Scientific, v.2.9).

Cryo-EM image processing and model building

All data processing was performed using cryoSPARC (v.3.2.0)⁵².

PhPINK1(D357A) dataset

Image stacks were corrected for beam-induced motion using patch motion, Fourier cropped to $0.833 \text{ \AA px}^{-1}$ and used for CTF estimation using patch CTF algorithm in cryoSPARC⁵². To create templates for particle picking, the micrographs were initially picked using blob picker, followed by particle extraction and 2D classification. Representative classes were used for template picking. Particles (310,371) were picked, extracted using a 380 px box size and processed for 2D classification. The best 235,948 particles were used to create an ab initio model. The 31,080 particles from the 2D classes, not resembling *PhPINK1*, were used to create a ‘junk’ ab initio model used as a trap for further data clean-up. Good particles were further cleaned using several rounds of heterogeneous refinement procedures first using the good and junk ab initio models followed by increasingly higher-resolution templates. The final set consisted of 216,021 particles and reached 2.48 \AA using homogeneous refinement protocol with imposed D_3 symmetry. To perform local refinement of the *PhPINK1* dimer, the symmetry was expanded to C_1 , yielding 1,295,406 particles. The *PhPINK1* dimer was modelled into the dodecamer map (see below), converted into a volume using UCSF Chimera⁵³, low passed to 12 \AA and used to create a soft padded mask in EMAN2 (54). This mask was used for local refinement of the *PhPINK1* dimer, yielding a 2.35 \AA map of the *PhPINK1* dimer.

Phosphorylated *PhPINK1* dataset

The initial data processing steps were similar to those for *PhPINK1(D357A)*. In brief, template picking yielded 205,887 particles and particle extraction was performed with

a 440 px box size at 0.78 Å px⁻¹. 2D classification yielded 139,028 particles, and heterogeneous refinement reduced the dataset to 89,061 particles. Homogeneous refinement protocol with imposed D_3 symmetry resulted in a 3.11 Å map. Symmetry expansion and local refinement yielded a 3.07 Å map of the phosphorylated *PhPINK1* dimer from 543,366 particles. To separate individual conformations of the α C helix in molecule B, we performed 3D variability analysis³⁴ in cluster mode with a mask around insertion-2, insertion-3 and the α C helix. Local refinement of cluster 1 (212,979 particles) using the dimer mask yielded a 3.25 Å map with a predominantly kinked α C helix. Local refinement of cluster 2 (172,764 particles) using the dimer mask yielded a 3.28 Å map with a predominantly extended α C helix. Cluster 3 (148,623 particles) represented *PhPINK1* states that could not be easily identified. All of the maps were processed for local-resolution estimation and local-resolution-based filtering using internal cryoSPARC (v.3.2.0) algorithms⁵².

Model building and refinement

The crystal structure of *PhPINK1* from the *PhPINK1*–Ub TVLN complex (PDB: [6EQI](#))¹⁰ was used as the initial model and was docked using UCSF Chimera (v.1.14)⁵³ into density corresponding to a monomer within the *PhPINK1*(D357A) dodecamer. Manual rebuilding of the model was performed in Coot (v.0.9)⁵⁰ and the additional N-helix included within the construct was built de novo. The model underwent multiple rounds of refinement using real space refine in Phenix (v.1.19.2-4158)⁴⁹. To generate a model of the dimer, the model was rebuilt and refined in one monomer, then duplicated and realigned into density of the opposing monomer, followed by further rounds of rebuilding and refinement. Multiplication of the *PhPINK1* dimer into the entire dodecamer map enabled generation of a dodecamer model. For dodecamer refinement, non-crystallography symmetry (NCS) restraints were used during refinement using real space refine, using two NCS groups corresponding to the two chains of the dimer. Owing to the low-resolution density corresponding to insertion-3 in the ‘kinked α C’ dimer (Fig. 4g and Extended Data Fig. 7c), an atomic model of insertion-3 could not be built de novo. As the shape of the density was reminiscent of the ordered insertion-3 from the *PhPINK1*–Ub TVLN complex¹⁰, insertion-3 (amino acids 259–280) was taken from the *PhPINK1*–Ub TVLN complex, rigid-body fitted into the density in Coot (v.0.9), stubbed at the C β carbons and merged with the rest of the model. This model was then passed once through real space refine in Phenix.

Data collection and refinement statistics are provided in Supplementary Table 2. All of the models were validated using MolProbity (v.4.5.1)⁵⁵. Structures were visualized and figures were generated in UCSF ChimeraX (v.1.1.1)⁵⁶.

In vitro H₂O₂ cross-linking assays

In vitro disulphide-linkage assays were performed at 22 °C by incubating 1.5 µM *PhPINK1* or *TcPINK1* with 2 mM H₂O₂ in 25 mM Tris (pH 7.4), 150 mM NaCl. At the indicated timepoints, the reaction was quenched in NuPAGE LDS sample buffer (Invitrogen), run on reducing or non-reducing NuPAGE 4–12% Bis-Tris gels (Invitrogen) and stained with InstantBlue Coomassie Protein Stain (Abcam).

In vitro kinase and oxidation assays

All *PhPINK1* autophosphorylation assays were performed in phosphorylation buffer (25 mM Tris (pH 7.4), 150 mM NaCl, 10 mM MgCl₂, 1 mM DTT) with 1.5 µM GST–*PhPINK1* (amino acids 115–575, WT, autophosphorylated) and 15 µM *PhPINK1* (amino acids 115–575) D334A or D357A (substrate). Reactions were started by the addition of 10 mM ATP and incubated at 22 °C for the indicated times.

For the simultaneous *PhPINK1* autophosphorylation and ubiquitin phosphorylation assay in Extended Data Fig. 6e, dephosphorylated *PhPINK1* (amino acids 119–575, oligomerization deficient) was first prepared by incubating 15 µM *PhPINK1* (amino acids 119–575) WT or S202A overnight at 4 °C with 7.5 µM λ-PP in SEC buffer supplemented with 2 mM MnCl₂, then purified by SEC. Kinase reactions were carried out as described above using 1.5 µM dephosphorylated *PhPINK1* (amino acids 119–575) and 15 µM ubiquitin.

Oxidation-coupled ubiquitin phosphorylation assays were performed in DTT-free phosphorylation buffer using 1.5 µM *PhPINK1* (amino acids 115–575) and 15 µM ubiquitin. *PhPINK1* was incubated with the indicated concentrations of H₂O₂ for 15 min at 4 °C, and the kinase reaction was initiated by the addition of 10 mM ATP and 15 µM ubiquitin and incubated at 22 °C for 3 h. To assess reversible *PhPINK1* oxidation, the indicated concentrations of H₂O₂ were added and incubated for 1 h at 4 °C, followed by the addition of 10 U ml⁻¹ catalase (Sigma-Aldrich) for 15 min to quench H₂O₂. The samples were divided and 10 mM DTT was added to one half to reverse *PhPINK1* oxidation. The kinase reaction was initiated by the addition of 10 mM ATP and 15 µM ubiquitin and incubated at 22 °C for 3 h.

All kinase reactions were quenched in SDS sample buffer (66 mM Tris (pH 6.8), 2% (w/v) SDS, 10% (v/v) glycerol, 0.005% (w/v) bromophenol blue) and run on custom-made 7.5% (for *PhPINK1* autophosphorylation) or 17.5% (for ubiquitin phosphorylation) Phos-tag gels (reducing) containing 50 µM Phos-tag Acrylamide AAL-107 (Wako) and 100 µM MnCl₂. Gels were stained with InstantBlue Coomassie Protein Stain (Abcam). Note that, in Phos-tag gels, phosphorylated proteins run markedly slower than unphosphorylated proteins, and no longer just according to their molecular mass. Therefore, size markers are not provided for the Phos-tag gels.

***Ph*PINK1 phosphosite identification by mass spectrometry**

*Ph*PINK1(D334A) phosphorylation was performed as described above, but with 10 mM DTT. The samples were run on SDS-PAGE and the band corresponding to *Ph*PINK1(D334A) was excised and destained twice with 50 mM ammonium bicarbonate/50% (v/v) acetonitrile. After dehydration with 100% (v/v) acetonitrile, the samples were reduced (10 mM TCEP for 30 min), alkylated (40 mM chloroacetamide for 30 min) and digested overnight (15 ng μ l⁻¹ of TPCK-treated trypsin at 37 °C). Peptides were extracted twice with 60% (v/v) acetonitrile/0.1% (v/v) formic acid and analysed on a timsTOFII pro mass spectrometer (Bruker) with PASEF-MS acquisition. Peptides were separated using a 90 min gradient (solvent A, 0.1% (v/v) formic acid; solvent B, 99.9% (v/v) acetonitrile/0.1% (v/v) formic acid) on a C18 analytical column (Aurora Series Emitter Column, AUR2-25075C18A 25 cm × 75 μ m × 1.6 μ m, IonOpticks). Data were searched using MaxQuant (v.1.6.17.0) at a 1% false-discovery rate, with oxidation and phosphorylation as variable modifications. The final MS2 spectra were reproduced in Skyline Daily (v.21.1.1.198)⁵⁷.

Analytical SEC to assess *Ph*PINK1 hetero-oligomerization

Analytical SEC was performed using a Superdex 200 Increase 3.2/300 column (Cytiva) equilibrated in SEC buffer. *Ph*PINK1 (amino acids 115–575) D334A, D357A and monomeric phosphorylated *Ph*PINK1 (amino acids 115–575, prepared as described above) were each, or in combination, diluted to 2 mg ml⁻¹ (per protein) in SEC buffer and incubated overnight at 4 °C. 50 μ l protein was loaded onto the column per run and eluted at 0.04 ml min⁻¹ flow rate.

Cell culture and constructs

HeLa *PINK1*^{-/-} cells were a gift from M. Lazarou (Monash University) and were authenticated at the Garvan Molecular Genetics facility using short tandem repeat profiling. Cells were cultured in DMEM supplemented with 10% (v/v) FBS (Gibco or Sigma-Aldrich), penicillin–streptomycin, and maintained at 37 °C and 5% CO₂. Cells were also screened routinely for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza). To limit the level of *Hs*PINK1 overexpression, the pcDNA5/FRT/TO plasmid was modified using the Q5 site-directed mutagenesis kit (NEB) to generate a 539 bp deletion in the CMV promoter (CMVd3)⁵⁸, hereafter referred to as pcDNA5^{d3}. The full-length, WT *Hs*PINK1 sequence was inserted into the BamHI site of pcDNA5^{d3} using In-Fusion Cloning (Takara Bio) and used for transient transfections. For stable *Hs*PINK1 expression, the *Hs*PINK1 sequence was inserted into the BamHI and NheI of the lentiviral pFU PGK

Hygro (pFUH) plasmid using InFusion Cloning. All of the *HsPINK1* mutants were generated using the Q5 site-directed mutagenesis kit (NEB).

Generation of stable cell lines

Stable YFP–Parkin and *HsPINK1*-expressing HeLa *PINK1*^{−/−} cell lines were generated using retroviral transduction of a pBMN-YFP-Parkin construct (gift from R. Youle; Addgene plasmid, 59416)⁵⁹ followed by lentiviral transduction of pFUH-*HsPINK1* WT or mutant constructs. For imaging, TOM20-Halo was incorporated by retroviral transduction of a pMIH-TOMM20-Halo construct (gift from B. Kile; Addgene plasmid, 111626)⁶⁰. All cells were selected by fluorescence sorting or by antibiotic selection.

Lattice light-sheet microscopy

Before imaging, cells were stained overnight using the JF646 HaloTag ligand according to the manufacturer’s instructions (Promega), then treated with 10 μM oligomycin and 4 μM antimycin A (OA) to depolarize mitochondria. Time-lapse live-cell data were acquired using a Lattice Light Sheet 7 (Zeiss, pre-serial). Light sheets (488 nm and 633 nm) of length 30 μm with a thickness of 1 μm were created at the sample plane using a ×13.3/0.44 NA objective. Fluorescence emission was collected via a ×44.83/1 NA detection objective. Aberration correction was set to a value of 182 to minimize aberrations as determined by imaging the Point Spread Function using 170 μm fluorescent microspheres at the coverslip of a glass-bottom chamber slide. Resolution was determined to be 454 nm (lateral) and 782 nm (axial). Data were collected with a range of frame rates of 16–20 ms and a z-step of 300 nm. Light was collected through a multi-band stop, LBF 405/488/561/633, filter. Images were collected at 37 °C and 5% CO₂.

Lattice light-sheet microscopy data analysis

Images taken using the Zeiss Lattice Light Sheet were processed in ZEN (Zeiss, v.3.5) using Zeiss’s deskew and deconvolution modules using a constrained iterative algorithm and 20 iterations. Maximum-intensity projections of the lattice videos were generated using Fiji (ImageJ 1.53k). To assess YFP–Parkin translocation, YFP–Parkin foci were manually counted. The time at which each cell began to show YFP–Parkin foci was recorded and the data were used to graph the cumulative fraction of cells that displayed YFP–Parkin foci over time. A linear fit was applied to the cumulative curves from the range of 20–80% of the translocated cells and used to calculate the time (min) at which 50% of cells exhibited YFP–Parkin translocation. A two-sided, two-sample Kolmogorov–Smirnov test was used to compare the cumulative distribution of

the two curves. A MATLAB script ([Supplementary Data](#)) was used to perform the Kolmogorov–Smirnov test. GraphPad Prism (v.9.0.0) was used to generate the graphs.

Transient transfection and western blotting

Cells were seeded in six-well plates one day before Lipofectamine 3000-mediated transfection of 1.5 µg pcDNA5^{d3}-*HsPINK1* WT or mutants. Then, 24 h after transfection, cells were treated with 10 µM oligomycin and 4 µM antimycin A (OA) for 2 h to depolarize mitochondria and induce *HsPINK1* accumulation. To induce the formation of disulphide-linked *HsPINK1* dimer, 2 mM H₂O₂ was added to the culture medium for 10 min before cell lysis. For stable YFP–Parkin/*HsPINK1*-expressing cell lines, cells were seeded two days before OA treatment and cell lysis. Cells were lysed directly in SDS sample buffer and run on reducing (or non-reducing for assessment of disulphide-linked PINK1 species) NuPAGE 4–12% Bis-Tris gels (Invitrogen). For Phos-tag western blots, the samples were run on custom-made 7.5% Phos-tag gels (reducing) containing 50 µM Phos-tag Acrylamide AAL-107 and 100 µM MnCl₂, and the gels were washed three times for 10 min in 10 mM EDTA followed by 10 min in water before transfer. Protein was transferred to PVDF membranes using the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked in 5% (w/v) milk powder in Tris-buffered saline containing 0.1% (v/v) Tween-20 (TBS-T) for 1 h, then incubated in primary antibodies (containing 3% (w/v) bovine serum albumin and 0.02% (w/v) sodium azide) overnight at 4 °C. The primary antibodies used were as follows: rabbit anti-PINK1 D8G3 1:1,000 (Cell Signaling Technology, 6946, 5), mouse anti-Parkin Prk8 1:1,000 (Cell Signaling Technology, 4211, 7), rabbit anti-phosphorylated ubiquitin (Ser65) 1:1,000 (Millipore, ABS1513-I, 3117322), rabbit anti-TOM20 FL-145 1:1,000 (Santa Cruz Biotechnology, sc-11415, D1613). Membranes were washed with TBS-T and incubated in goat anti-rabbit HRP-conjugated secondary antibodies (1:5,000, SouthernBiotech, 4010-05, A4311-TF99D) or goat anti-mouse HRP-conjugated secondary antibodies (1:5,000, SouthernBiotech, 1030-05, E2518-Z929D) for 1 h at room temperature before washing with TBS-T and detection using the ChemiDoc (Bio-Rad) after HRP substrate incubation. For tubulin blots, the membranes were incubated in hFAB rhodamine anti-tubulin antibodies (1:5,000; Bio-Rad, 12004165) for 1 h at room temperature or overnight at 4 °C followed by washing in TBS-T and detection using the ChemiDoc (Bio-Rad).

Subcellular fractionation and ubiquitin phosphorylation assay

HeLa *PINK1*^{−/−} cells (1 × 10⁶) were seeded in 10 cm dishes. The next day, cells were transiently transfected with 5 µg pcDNA5^{d3}-*HsPINK1* WT or C166S/S167N (SN) and, the next day, treated with OA for 2 h to accumulate *HsPINK1*. Cells were permeabilized in 1 ml fractionation buffer (20 mM HEPES (pH 7.4), 250 mM sucrose,

50 mM KCl, 2.5 mM MgCl₂) supplemented with 0.025% (w/v) digitonin, protease inhibitor cocktail tablet (Roche) and PhosSTOP (Roche) for 20 min at 4 °C. Heavy membranes were pelleted by centrifugation for 5 min at 14,000g at 4 °C, followed by washing with fractionation buffer. Heavy membrane pellets were resuspended in 100 µl fractionation buffer and divided into 50 µl fractions. Each fraction was incubated with the indicated concentrations of H₂O₂ with gentle agitation for 30 min at 4 °C. Catalase (10 U ml⁻¹; Sigma-Aldrich) was added for 15 min to quench H₂O₂, and 10 mM DTT was added to select fractions for 15 min to study reversible oxidation. To initiate the kinase reaction, 10 mM ATP and 15 µM ubiquitin was added and the mixture was incubated 1 h at 30 °C with agitation. Heavy membranes (containing *HsPINK1*) and supernatant (containing Ub) were separated by centrifugation and samples of both were prepared in SDS sample buffer. Western blotting was performed as described above. ImageLab (Bio-Rad, v.6.1) was used to quantify band intensities for the phosphorylated monoubiquitin band, from three independent experiments (Fig. 5e and Supplementary Fig. 2).

AlphaFold structure prediction

The structure of *HsPINK1* as predicted by AlphaFold2 was obtained from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>)^{37,38}. To predict the structure of Parkin Ubl-bound *HsPINK1* and dimerized *HsPINK1*, we used the ColabFold Google Colab notebook called AlphaFold2_complexes³⁹. The predicted model with the highest IDDT score is shown.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The coordinates and crystallographic structure factors for *PhPINK1*(D334A) have been deposited at the PDB under accession code [7T3X](#), and EM models and maps under accession codes [7T4M](#) (*PhPINK1*(D357A) dodecamer; Electron Microscopy Data Bank (EMDB): [EMD-25680](#)), [7T4N](#) (*PhPINK1*(D357A) dimer; EMDB: [EMD-25681](#)), [7T4L](#) (WT phosphorylated *PhPINK1* dimer with extended αC helix in chain B; EMDB: [EMD-25679](#)) and [7T4K](#) (WT phosphorylated *PhPINK1* dimer with kinked αC helix in chain B; EMDB: [EMD-25678](#)). We also deposited the map for WT phosphorylated *PhPINK1* dimer before 3D variability analysis (EMDB: [EMD-25677](#)). Uncropped versions of all gels and blots are provided in Supplementary Fig. 1. [Source data](#) are provided with this paper.

Code availability

The MATLAB script that was used to perform the Kolmogorov–Smirnov test to analyse lattice light-sheet imaging data is provided in the [Supplementary Data](#).

References

1. Valente, E. M. et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158–1160 (2004).
2. Pickrell, A. M. & Youle, R. J. The roles of PINK1, Parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* **85**, 257–273 (2015).
3. Kondapalli, C. et al. PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating serine 65. *Open Biol.* **2**, 120080 (2012).
4. Kazlauskaite, A. et al. Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65. *Biochem. J.* **460**, 127–139 (2014).
5. Koyano, F. et al. Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* **510**, 162–166 (2014).
6. Kane, L. A. et al. PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *J. Cell Biol.* **205**, 143–153 (2014).
7. Ordureau, A. et al. Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Mol. Cell* **56**, 360–375 (2014).
8. Wauer, T. et al. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J.* **34**, 307–325 (2015).

9. Gladkova, C. et al. An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. *EMBO J.* **36**, 3555–3572 (2017).
10. Schubert, A. F. et al. Structure of PINK1 in complex with its substrate ubiquitin. *Nature* **552**, 51–56 (2017).
11. Kumar, A. et al. Structure of PINK1 and mechanisms of Parkinson’s disease-associated mutations. *eLife* **6**, e29985 (2017).
12. Okatsu, K. et al. Structural insights into ubiquitin phosphorylation by PINK1. *Sci. Rep.* **8**, 10382 (2018).
13. Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression and mortality. *Neurology* **17**, 427–442 (1967).
14. Corti, O., Lesage, S. & Brice, A. What genetics tells us about the causes and mechanisms of Parkinson’s disease. *Physiol. Rev.* **91**, 1161–1218 (2011).
15. Nguyen, T. N., Padman, B. S. & Lazarou, M. Deciphering the molecular signals of PINK1/Parkin mitophagy. *Trends Cell Biol.* **26**, 733–744 (2016).
16. Harper, J. W., Ordureau, A. & Heo, J.-M. Building and decoding ubiquitin chains for mitophagy. *Nat. Rev. Mol. Cell Biol.* **19**, 93–108 (2018).
17. Lazarou, M., Jin, S. M., Kane, L. A. & Youle, R. J. Role of PINK1 binding to the TOM complex and alternate intracellular membranes in recruitment and activation of the E3 ligase Parkin. *Dev. Cell* **22**, 320–333 (2012).
18. Okatsu, K. et al. A dimeric PINK1-containing complex on depolarized mitochondria stimulates Parkin recruitment. *J. Biol. Chem.* **288**, 36372–36384 (2013).
19. Wauer, T., Simicek, M., Schubert, A. F. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **524**, 370–

374 (2015).

20. Kumar, A. et al. Parkin–phosphoubiquitin complex reveals cryptic ubiquitin-binding site required for RBR ligase activity. *Nat. Struct. Mol. Biol.* **24**, 475–483 (2017).
21. Gladkova, C., Maslen, S. L., Skehel, J. M. & Komander, D. Mechanism of parkin activation by PINK1. *Nature* **559**, 410–414 (2018).
22. Sauvé, V. et al. Mechanism of parkin activation by phosphorylation. *Nat. Struct. Mol. Biol.* **25**, 623–630 (2018).
23. Woodroof, H. I. et al. Discovery of catalytically active orthologues of the Parkinson’s disease kinase PINK1: analysis of substrate specificity and impact of mutations. *Open Biol.* **1**, 110012 (2011).
24. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
25. Patel, O. et al. Structure of SgK223 pseudokinase reveals novel mechanisms of homotypic and heterotypic association. *Nat. Commun.* **8**, 1157 (2017).
26. Ha, B. H. & Boggon, T. J. The crystal structure of pseudokinase PEAK1 (Sugen kinase 269) reveals an unusual catalytic cleft and a novel mode of kinase fold dimerization. *J. Biol. Chem.* **293**, 1642–1650 (2018).
27. Sekine, S. et al. Reciprocal roles of Tom7 and OMA1 during mitochondrial import and activation of PINK1. *Mol. Cell* **73**, 1028–1043 (2019).
28. Hasson, S. A. et al. High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291–295 (2013).

29. Rasool, S. et al. Mechanism of PINK1 activation by autophosphorylation and insights into assembly on the TOM complex. *Mol. Cell* **82**, 44–59 (2022).
30. Rasool, S. et al. PINK1 autophosphorylation is required for ubiquitin recognition. *EMBO Rep.* **19**, e44981 (2018).
31. Aerts, L., Craessaerts, K., Strooper, B. D. & Morais, V. A. PINK1 kinase catalytic activity is regulated by phosphorylation on serines 228 and 402. *J. Biol. Chem.* **290**, 2798–2811 (2015).
32. Aerts, L., Craessaerts, K., Strooper, B. D. & Morais, V. A. In vitro comparison of the activity requirements and substrate specificity of human and triboleum castaneum PINK1 orthologues. *PLoS ONE* **11**, e0146083 (2016).
33. Okatsu, K. et al. PINK1 autophosphorylation upon membrane potential dissipation is essential for Parkin recruitment to damaged mitochondria. *Nat. Commun.* **3**, 1016 (2012).
34. Punjani, A. & Fleet, D. J. 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
35. Soylu, İ. & Marino, S. M. Cy-preds: an algorithm and a web service for the analysis and prediction of cysteine reactivity. *Proteins Struct. Funct. Bioinform.* **84**, 278–291 (2016).
36. Shrestha, S. et al. A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase superfamily. *Sci. Signal.* **13**, eaax6313 (2020).
37. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
38. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

39. Ovchinnikov, S., Mirdita, M. & Steinegger, M. ColabFold—making protein folding accessible to all via Google Colab (v1.0-alpha). *Zenodo* <https://doi.org/10.5281/zenodo.5123297> (2021).
40. Pollock, L., Jardine, J., Urbé, S. & Clague, M. J. The PINK1 repertoire: not just a one trick pony. *Bioessays* **43**, 2100168 (2021).
41. Hertz, N. T. et al. A neo-substrate that amplifies catalytic activity of Parkinson’s-disease-related kinase PINK1. *Cell* **154**, 737–747 (2013).
42. Lowe, E. D. et al. The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition. *EMBO J.* **16**, 6646–6658 (1997).
43. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
44. Aragão, D. et al. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* **25**, 885–891 (2018).
45. McPhillips, T. M. et al. Blu-Ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 (2002).
46. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
47. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
48. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
49. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
51. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
52. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
53. Yang, Z. et al. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J. Struct. Biol.* **179**, 269–278 (2012).
54. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
55. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
56. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
57. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
58. Slater, M. et al. Achieve the protein expression level you need with the mammalian HaloTag 7 flexi vectors. *Promega Notes* **100**, 16–18 (2008).
59. Yamano, K., Fogel, A. I., Wang, C., Bliek, A. M. V. D. & Youle, R. J. Mitochondrial Rab GAPs govern autophagosome biogenesis during mitophagy. *eLife* **3**, e01612 (2014).
60. McArthur, K. et al. BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. *Science* **359**, eaao6047 (2018).

61. Wang, W. et al. Atomic structure of human TOM core complex. *Cell Discov.* **6**, 67 (2020).

Acknowledgements

We thank the members of the Ubiquitin Signalling Division at WEHI and I. Lucet (WEHI) and M. Lazarou (Monash) for reagents, advice and comments on the manuscript; J. Newman (CSIRO C3 Crystallisation Centre) for help with crystallization; and the staff at the WEHI Information Technology Services and the WEHI Research Computing Platform for providing facilities and support. This work was funded by WEHI, an NHMRC Investigator Grant (GNT1178122 to D.K.), the Michael J. Fox Foundation and Shake-It-Up Australia (to D.K. and G.D.). A.G. is a CSL Centenary Fellow. G.D. is supported by a fellowship from the Bodhi Education Fund and NHMRC Ideas grant (GNT2004446). Z.Y.G. is supported by an Australian Government Research Training Program Scholarship. Research was further supported by an NHMRC Independent Research Institutes Infrastructure Support Scheme grant (361646) and Victorian State Government Operational Infrastructure Support grant. This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. We also acknowledge use of the Titan Krios and other facilities at the Ian Holmes Imaging Centre at the Bio21 Molecular Science and Biotechnology Institute.

Author information

Affiliations

1. Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

Zhong Yan Gan, Sylvie Callegari, Simon A. Cobbold, Thomas R. Cotton, Michael J. Mlodzianoski, Niall D. Geoghegan, Kelly L. Rogers, Andrew Leis, Grant Dewson, Alisa Glukhova & David Komander

2. Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia

Zhong Yan Gan, Sylvie Callegari, Simon A. Cobbold, Thomas R. Cotton, Michael J. Mlodzianoski, Niall D. Geoghegan, Kelly L. Rogers, Andrew Leis, Grant Dewson, Alisa Glukhova & David Komander

3. Medical Research Council Laboratory of Molecular Biology, Cambridge, UK

Alexander F. Schubert

4. Drug Discovery Biology, Monash Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia

Alisa Glukhova

5. Department of Biochemistry and Pharmacology, University of Melbourne, Melbourne, Victoria, Australia

Alisa Glukhova

Contributions

D.K. conceived the project. Z.Y.G. designed and performed all experiments. S.C. and G.D. contributed to the design of cell biology experiments. D.K. and T.R.C. contributed to crystallography. S.A.C. performed mass spectrometry experiments. T.R.C. performed SEC–MALS experiments. S.C., M.J.M., N.D.G. and K.L.R. performed and analysed lattice light sheet microscopy experiments. A.F.S. contributed ideas, advice and preliminary data. A.L. and A.G. performed EM data collection, A.G. performed data processing and map calculations and Z.Y.G. performed model building. D.K. and Z.Y.G. analysed the data and wrote the manuscript with contributions from all of the authors.

Corresponding author

Correspondence to [David Komander](#).

Ethics declarations

Competing interests

D.K. serves on the Scientific Advisory Board of BioTheryX Inc. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Tony Hunter, Titia Sixma and Richard Youle for their contribution to the peer review of this work. Peer reviewer reports are available.

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 Crystal structures of unphosphorylated PINK1.](#)

a, Constructs used in crystal structures of *Tc*PINK1 (PDB: 5OAT¹¹ and 5YJ9¹²). Phosphomimetics substituted for Ser205 (equivalent to Ser202 in *Ph*PINK1) and other residues, although one structure (PDB: 5OAT) still showed phosphorylation at Ser207 (equivalent to Ser204 in *Ph*PINK1, and Ser230 in *Hs*PINK1, see below). Constructs were further engineered as indicated and described^{11,12}. **b**, Depiction of *Tc*PINK1 crystal structures, with key features indicated. **c**, $2|F_o| - |F_c|$ electron density contoured at 1.5σ ,

covering the *Ph*PINK1(D334A) molecule in the asymmetric unit. Right, detail for regions of interest including the α C helix and Ser202-containing loop and the N-helix. **d**, Our crystal structure of unphosphorylated *Ph*PINK1(D334A) (compare Fig. 1) in the same orientation as in **b**, for comparison. **e**, Two different zoomed views of the ATP-binding site of unphosphorylated *Ph*PINK1, with an ATP molecule modelled in from PDB 2PHK (ref. 42) as before¹⁰. Left, highlighting key features common to most active kinases, including the P-loop (grey), α C helix (orange) and activation segment (pink). The DFG and HRD motifs (dotted outlines) which include Asp357 involved in coordination of Mg²⁺, and the substrate-binding Asp334 involved in phosphoryl transfer (and mutated to Ala in the crystal structure) are indicated. Glu214 and Lys193 form a crucial salt bridge to coordinate ATP. Right, assembled catalytic (C) and regulatory (R) spines²⁴ in red and blue, respectively, indicative of an active kinase conformation. **f**, Crystal contact involving N-helix residue Trp129, which binds to the N-lobe of a symmetry related molecule. ASU, asymmetric unit.

Extended Data Fig. 2 The N-helix and its role in keeping PINK1 monomeric.

a, The pseudokinases SgK223 (PDB: 5VE6) and SgK269 (PDB: 6BHC) were structurally characterised only recently^{25,26}, and contain a CTR similar to PINK1 that is complemented by an N-helix, in a highly analogous fashion now shown for PINK1. While conceptually similar, the organization of the N-helix on the CTR is however distinct. SgK223 and SgK269 utilize a cross-architecture to provide a dimerization interface, whereas the orientation in PINK1 is parallel to the α K helix of the CTR. **b**, SEC–MALS analysis of *Ph*PINK1 with (amino acids 115–575) or without (amino acids 143–575) the N-helix. The shorter *Ph*PINK1 construct tends to form less well-defined dimers, whereas the longer variant is a monomer. Experiments were performed three times with identical results. **c**, Previous *Tc*PINK1 and *Ph*PINK1 structures dimerized in the crystal lattice through the CTR, reflecting an available interaction surface since the N-helix was missing. *Tc*PINK1 structures dimerized identically (left, only shown for PDB 5OAT¹¹). The relative orientation of *Ph*PINK1 molecules in PDB 6EQI¹⁰ was different to *Tc*PINK1 (middle). Right, *Ph*PINK1(D334A) structure

including the N-helix, for comparison. **d**, Schematic of the N-helix–CTR interaction, and situation in previous PINK1 structures, as identified in structures and in SEC–MALS.

Extended Data Fig. 3 PhPINK1(D357A) oligomerization enables EM studies.

a, Elution profile of the *PhPINK1(D357A)* mutant during purification by SEC. Data shown is from a representative experiment of three runs. **b**, Thermal denaturation studies of purified *PhPINK1* (amino acids 115–575) variants. The crystallized D334A mutant has a melting curve profile and melting temperature similar to WT *PhPINK1*. *PhPINK1(D357A)* shows an unusual profile with a high secondary melting temperature. Technical duplicates were measured in three independent experiments. Average melting temperatures are indicated. **c**, Negative stain EM analysis of *PhPINK1(D357A)* revealed a highly ordered oligomer suitable for cryo-EM studies. Minimal processing was performed in RELION (v.3.1)⁵¹, and a subset of the resulting 2D classes is depicted as insets. Scale bar, 50 nm. Negative staining for this sample was performed once, before advancing to cryo-EM analysis. **d**, Flowchart for cryo-EM analysis as described in [Methods](#). **e**, 2D classification of particles reveal a 2D-crystalline arrangement of *PhPINK1(D357A)* oligomers in some areas of the grid. **f**, Final cryo-EM density maps (coloured by local resolution) for the *PhPINK1(D357A)* dodecamer (left) at 2.48 Å and the extracted dimer (right) at 2.35 Å. **g**, Examples of map quality for the final 2.35 Å density covering the *PhPINK1(D357A)* dimer.

[Source data](#)

Extended Data Fig. 4 Further analysis of the *PhPINK1* oligomeric state.

a, Model built into the *PhPINK1(D357A)* dodecamer cryo-EM density as in Fig. [2b](#). **b**, Side view of the oligomer showing how the N-helix–CTR area facilitates contacts to connect four dimers. **c**, We were concerned that *PhPINK1* oligomerization may only happen with a specific mutant protein and tested whether the oligomer would also form with unphosphorylated

WT *PhPINK1*. WT *PhPINK1* is phosphorylated at multiple sites when purified from *E. coli* (see¹⁰), and was hence dephosphorylated by λ-PP and repurified by SEC as indicated (see [Methods](#)). Like the inactive D357A mutant, a prominent oligomer, as well as a dimer/monomer equilibrium is apparent on SEC. Alternatively, the WT protein was first dephosphorylated by λ-PP, then rephosphorylated by adding Mg²⁺/ATP for 1 min, and the kinase and λ-PP were inactivated by EDTA (see [Methods](#)).

Phosphorylation destabilized the oligomer to a predominant monomeric species, explaining why WT *PhPINK1* does not form the oligomer, and suggesting that autophosphorylation resolves the oligomer and dimer (see below). These experiments were part of the purification process and were performed three times with identical results. **d**, SEC–MALS analysis of *TcPINK1* variants, to show that corresponding *TcPINK1* constructs (amino acids 117–570) do not show oligomeric behaviour. *TcPINK1* purifications revealing monomeric behaviour were performed at least twice, and a SEC–MALS experiment was performed once. **e**, Residues mediating hydrogen bonds (dotted lines) in the oligomer interfaces. Lack of conservation of oligomer contacts between *PhPINK1*, *TcPINK1* and *HsPINK1* likely explain why *TcPINK1* does not form an oligomer, and why we do not expect an identical oligomer in *HsPINK1*.

[Extended Data Fig. 5 Structural and biochemical analysis of the *PhPINK1* dimer.](#)

a, Comparison between *PhPINK1* autophosphorylation and ubiquitin phosphorylation resolved previously (PDB: 6EQI)¹⁰, with relevant details in the active site (top row) and in the activation segment (bottom row). The right panel shows substrate disposition in relation to a modelled ATP molecule as in Fig. [3b](#). The dimeric *PhPINK1* autophosphorylation complex appears to place Ser202 in an ideal phospho-accepting position. **b**, Comparison of activation segment structures in all published PINK1 structures, revealing high structural similarity. Ser375 (*PhPINK1*)/Ser377 (*TcPINK1*) is shown in ball-and-stick representation; this residue was mutated to Asp in one prior structure (see Extended Data Fig. [1a](#)). **c**, *PhPINK1* Ser375 corresponds to Ser402 in *HsPINK1*, which is a reported phosphorylation site^{31,33}, but is located within the activation segment and

out of reach of the substrate-binding site within dimeric *PhPINK1*. Our structure does not reveal how autophosphorylation at this residue could be facilitated in *cis* or *trans* (left), nor how phosphorylation through e.g. an upstream kinase would contribute to PINK1 activity or function, since phosphorylation would likely disrupt the dimer (right). ATP was modelled as in Fig. 3b. **d**, Manual docking of dimeric *PhPINK1* onto a cryo-EM structure of dimeric human TOM complex (PDB: 7CK6)⁶¹. The *PhPINK1* dimer was oriented with its two N-helices (spanning ~80 Å) aligning with the two TOM7 subunits of the TOM complex dimer. TOM7 has been reported as essential for PINK1 stabilization on the TOM complex^{27,28}. Note that some TOM components have considerable cytosolic domains that would need to be accommodated in addition to a PINK1 dimer. **e**, Unidentified density in the 2.35 Å cryo-EM map connects Cys169 in the dimer. **f**, Time course of *PhPINK1* and *TcPINK1* disulphide formation upon treatment with 2 mM H₂O₂, resolved on a non-reducing SDS-PAGE gel. The dimer/oligomer stabilizing *PhPINK1* D357A mutation (Fig. 3) enables fast disulphide formation that is averted by an additional C169A mutation. *TcPINK1* WT or D359A (equivalent to *PhPINK1*(D357A)) mutant do not show fast cross-linking behaviour observed in *PhPINK1*. Engineering of an additional Thr172 to Cys mutation (*TcPINK1*(T172C/D359A)) results in rapid emergence of cross-linked *TcPINK1* dimers upon oxidation. Experiments were performed in biological triplicate with identical results. See Supplementary Fig. 1 for uncropped gels. **g**, EOPD mutations in the activation segment and αEF–αF loop. Mutations according to¹⁰. **h**, *HsPINK1* EOPD mutants listed in **g** were expressed in HeLa *PINK1*^{-/-} cells, stabilized with OA and treated with H₂O₂ to assess PINK1 dimerization, autophosphorylation and ubiquitin phosphorylation activity (see [Methods](#)). The control D384A mutant (mutation of the DFG motif, equivalent to D357A in *PhPINK1*) was included as an inactive mutant. As anticipated, it is able to dimerize (oxidative cross-link formed) but unable to autophosphorylate or generate phosphorylated ubiquitin. Oxidative dimerization enables separation of pure catalytic mutants and dimerization deficient mutants. Experiments were performed in biological triplicate with identical results. See Supplementary Fig. 1 for uncropped blots.

Extended Data Fig. 6 Analysis of autophosphorylation sites in PhPINK1 and HsPINK1.

a, Phos-tag analysis of inactive, monomeric *PhPINK1*(D334A), which does not autophosphorylate when incubated with ATP. Phosphorylation by WT GST-tagged *PhPINK1* for 2 h leads to a shift of the entire protein consistent with a single phosphorylation site, and a small amount (1–2%) shifts to a higher species indicating additional autophosphorylation events after prolonged incubation. A representative gel of three independent experiments is shown. See Supplementary Fig. 1 for uncropped gel. **b**, As in **a** but shown in a time course experiment. Mutating Ser202 to Ala abrogates the observed gel shift, indicating that Ser202 is the sole site of phosphorylation. Mutation of Ser375 (equivalent to *HsPINK1* Ser402) to Ala does not impede phosphorylation. A representative gel of three independent experiments is shown. See Supplementary Fig. 1 for uncropped gel. **c**, Mass spectrometry confirms the detection of a phosphate at Ser202 of *PhPINK1*(D334A) after phosphorylation with WT GST–*PhPINK1*. See [Methods](#). **d**, Time course experiment as in **b**, with monomeric *PhPINK1*(D334A) or oligomeric *PhPINK1*(D357A). While the monomer can be phosphorylated by WT GST–*PhPINK1*, the oligomer is not efficiently phosphorylated since Ser202 is buried in the stable dimer/oligomer structure. This experiment was performed twice with identical results. See Supplementary Fig. 1 for uncropped gel. **e**, Time course experiment using dephosphorylated WT *PhPINK1* (amino acids 119–575, which we found to be oligomerization impaired) and *PhPINK1* (amino acids 119–575) S202A, and with ubiquitin in the reaction. See [Methods](#) for details. Phosphorylation of *PhPINK1* at Ser202 is essential for ubiquitin phosphorylation, as both phosphorylation events are completely abrogated if Ser202 cannot be phosphorylated. Representative gels of three independent experiments are shown. See Supplementary Fig. 1 for uncropped gels. **f–i**, Analysis of autophosphorylation sites in *HsPINK1*. **f**, In *HsPINK1*, the Ser202 equivalent residue is Ser228, which is a known, important autophosphorylation site. In *PhPINK1*, the Ser202 site is followed by Asn203 and Ser204, and Ser204 can be phosphorylated when *PhPINK1* is expressed in *E. coli*¹⁰. Also, the Ser204 equivalent residue in *TcPINK1* (Ser207) was phosphorylated in a previous crystal structure (Extended Data Fig. 1a). In *HsPINK1*, Ser228 is followed by Ser229 and

Ser230, and we investigated whether phosphorylation of these residues occurs and whether it contributes to *HsPINK1* activity. **(g)** Structural detail of the Ser202-containing loop in the active site of the *PhPINK1* autophosphorylation dimer. Equivalent *HsPINK1* residues are in brackets. ATP was modelled as in Fig. 3b. Only Ser202 is in the phospho-acceptor site, but the other residues may occupy the site and be phosphorylated subsequently, if the end of the α C helix slightly unravels; since we expect conformational changes in this region (see below), this seemed feasible. **h**, **i**, *HsPINK1* variants with combinations in the Ser228-containing loop as indicated (empty, vector control; KD, kinase dead, a triple mutant K219A, D362A (HRD motif), D384A (DFG motif); SSS (WT) refers to the WT sequence with Ser228, Ser229, Ser230; ASS refers to Ala228, Ser229, Ser230; etc.) were expressed in HeLa *PINK1*^{-/-} cells, treated with OA for 2 h, and subjected to western blotting (see [Methods](#)). **h**, A Phos-tag gel probed with anti-PINK1 antibody reveals that KD and AAA mutants remain unphosphorylated, while the WT *HsPINK1* (SSS) shows multiple phosphorylation states, indicating more than one phosphorylation event in this loop. All three Ser residues appear to be phosphorylatable. These results were correlated with appearance of Ser65-phosphorylated ubiquitin in the same cell lysates, revealed by a ubiquitin Ser65-phosphospecific antibody. WT *HsPINK1* leads to a strong signal for phosphorylated ubiquitin, while Ala mutation in Ser228 does not lead to a phosphorylated ubiquitin signal. In contrast, if only Ser228 is present (e.g. SAA mutant) the phosphorylated ubiquitin signal is indistinguishable from that of WT *HsPINK1*. **i**, As in **h**, but with further substitution of Ser228. Phosphomimetic residues Asp228 or Glu228, followed by Ala229 and Ala230, lead to a small increase in phosphorylated ubiquitin, in the absence of autophosphorylation. The results in **h** and **i** indicate that also in *HsPINK1*, Ser228 phosphorylation is essential to turn PINK1 into a ubiquitin kinase, and a phosphomimetic is a weak substitute. Other residues in this area may also become phosphorylated but do not trigger ubiquitin phosphorylation. Experiments shown in **h** and **i** were performed in biological triplicate with identical results. See Supplementary Fig. 1 for uncropped blots.

Extended Data Fig. 7 Structure of a cross-linked, phosphorylated *PhPINK1* reveals asymmetry and

conformational changes in the N-lobe.

a, Final phosphorylated and cross-linked *PhPINK1* oligomer on a non-reducing SDS–PAGE gel. See Supplementary Fig. 1 for uncropped gel. **b**, Flowchart for cryo-EM analysis, as in Extended Data Fig. 3. **c**, Local resolution maps and resolution calculations. **d**, Map quality for selected regions. **e**, EM density for the phosphorylated Ser202 of molecule A, in the substrate-binding site of molecule B.

Extended Data Fig. 8 Assessment of *PhPINK1* dimer stability by measuring oligomer formation.

a, Our studies had revealed that disruption of the kinase–substrate interaction had a detrimental effect for dimer formation. We wondered whether the underlying dimerization mechanism would facilitate the reported effect that each copy of PINK1 is phosphorylated during mitophagy^{18,30,33}. Based on our biochemical data (Fig. 2a and Extended Data Fig. 4c) we conceptualized PINK1 dimer interactions as a function of the number of intact Ser202–Asp334 contacts. Ser202 phosphorylation would resolve the Ser202–Asp334 contacts individually, leading to one contact point after the first phosphorylation event, and zero contact points when both PINK1 molecules are phosphorylated. It was unclear whether the dimer is stable after the first phosphorylation event, when only one Ser202–Asp334 contact point exists within the dimer. **b**, We explored whether we would be able to generate *PhPINK1* oligomers with an unphosphorylated *PhPINK1*(D334A) mutant and Ser202-phosphorylated *PhPINK1* (phospho-*PhPINK1*), as indicated, with the knowledge that homo-oligomerization is disfavoured (zero Ser202–Asp334 contact points, highlighted in red). Formation of a hetero-oligomer between unphosphorylated *PhPINK1*(D334A) and phospho-*PhPINK1* would indicate that the dimer is stable with just one Ser202–Asp334 contact (highlighted in green). **c**, SEC analysis of the mutants confirms that unphosphorylated *PhPINK1*(D334A) or phospho-*PhPINK1* do not form oligomers, and hence, zero Ser202–Asp334 contact points are not sufficient. In contrast, enabling one contact point, as per the green scenario in **b** where variants are mixed, enables oligomerization of *PhPINK1*. Each individual protein showed identical results in every purification run ($n = 3$),

and the mixing experiments were performed in biological triplicate. **d**, Phos-tag analysis confirms that the oligomer species is a 1:1 mixture of unphosphorylated *PhPINK1*(D334A) and phospho-*PhPINK1*. Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped gel. Together, these experiments show that the dynamic equilibrium favours dimer formation until each copy of PINK1 has been phosphorylated, ensuring full phosphorylation.

Extended Data Fig. 9 Regulation of PINK1 activity by oxidation.

a, Extended sequence alignment indicating that Cys169 and Cys360 are well conserved in PINK1, and invariant in mammalian PINK1. Cys169 is a Thr in *TcPINK1*, and a Ser in many fish species. **b**, Comparison of *PhPINK1* and *TcPINK1*, for their ability to be regulated by oxidation, shown in Phos-tag ubiquitin phosphorylation assays (see Fig. 5b). While *PhPINK1* activity is abrogated with H₂O₂, *TcPINK1* with Thr172 in the P-loop, remains active. The observed reduction in *TcPINK1* activity could be a result of oxidation of the conserved Cys362 (Cys360 in *PhPINK1*) in the active site. Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped gel. **c**, Inhibition of *PhPINK1* ubiquitin phosphorylation activity can be reversed with DTT, suggesting reversible regulatory oxidation. See [Methods](#). Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped gel. **d**, Time course assessment of *HsPINK1* Cys–Ala mutants transiently expressed in HeLa *PINK1*^{−/−} cells. OA treatment leads to accumulation of *HsPINK1* and slightly altered autophosphorylation for both C166A and C387A. The *HsPINK1*(C166A) mutant was almost completely deficient in ubiquitin phosphorylation, while *HsPINK1*(C387A) showed highly reduced but still detectable phosphorylated ubiquitin levels. Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped blots. **e**, Time course of *HsPINK1* mutants as in **d**, but using stable *HsPINK1* expression in the presence of YFP–Parkin. Presence of YFP–Parkin seemingly increases levels of phosphorylated ubiquitin for all *HsPINK1* variants as compared to **d**, yet overall, phosphorylated ubiquitin and phosphorylated Parkin levels remain strongly diminished with *HsPINK1*(C166A), and to a lesser degree with *HsPINK1*(C387A).

Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped blots. **f**, Translocation of YFP–Parkin (cyan) to mitochondria (magenta, TOM20–Halo) in HeLa *PINK1*^{-/-} stably expressing *HsPINK1* Cys–Ala variants upon OA treatment, imaged using lattice light sheet microscopy. Maximum intensity projections are shown for four different timepoints. YFP–Parkin translocation is delayed in cells expressing either *HsPINK1*(C166A) or *HsPINK1*(C387A) relative to WT h*HsPINK1*. A kinase dead (KD) *HsPINK1* variant was included as a control. Images are representative of three independent experiments. Scale bar is 10 μm. See Supplementary Video 1. **g**, Quantification of YFP–Parkin translocation in **f**. The cumulative fraction of cells exhibiting YFP–Parkin translocation is shown over time. Approximately 70 cells were counted per cell line. Each curve was fitted to determine the time for 50% of the cells to feature translocation. Significant differences between curves were determined using a two-sample Kolmogorov–Smirnov test. A MATLAB script and Source Data are available as [Supplementary Material](#). Exact *p* values are: WT–C166A: *p* < 2.83×10^{-20} ; WT–C387A: *p* < 2.36×10^{-18} . **h**, *HsPINK1* mutants expressed in HeLa *PINK1*^{-/-} cells and treated with OA for 2 h. Activity of *HsPINK1*(C166A) and *HsPINK1*(C166S) can be partially restored by an additional S167N mutation, mimicking the sequence observed in many fish species. A *TcPINK1*-like sequence introduced into *HsPINK1*, C166T/S167N, is less active than the *HsPINK1*(C166S/S167N) mutant. The additional Asn in the P-loop of the *HsPINK1*(C166S/S167N) mutant recovers activity of *HsPINK1*(C166S), suggesting that both residues are also important in ubiquitin/Ubl substrate interactions, and not merely involved in dimerization. Experiments were performed in biological triplicate. See Supplementary Fig. 1 for uncropped blots. **i**, A redox active switch in fructosamine-3-kinases (PDB: 6OID) is conceptually similar, utilizing a Cys at an identical position (Cys32) for Cys-mediated cross-linking and regulation of kinase activity by oxidation³⁶, however the overall orientation of kinase domains is dissimilar.

[Source data](#)

[**Extended Data Fig. 10 Analysis of the *HsPINK1* model as predicted by AlphaFold2**](#)^{37,38}.

(a, b) Predicting the structure of *HsPINK1* via AlphaFold2 (**a**, left), resulted in a model remarkably similar to *PhPINK1* from the *PhPINK1–Ub* TVLN complex¹⁰ (**b**, right), and features a kinked α C helix, ordered insertion-3, and an extended N-helix that binds and extends from the CTR directly into the membrane. Consistently, predicting a complex between *HsPINK1* with the ubiquitin-like (Ubl) domain of human Parkin (**a**, right), places the Ubl domain at *HsPINK1* analogously to the *PhPINK1–Ub* TVLN complex (**b**, right). Both predictions are dissimilar to unphosphorylated *PhPINK1* (**b**, left) and *TcPINK1* (Extended Data Fig. 1). Insets show the detail of the N-lobe with a kinked α C helix and an ordered insertion-3. The prediction is somewhat surprising, as it suggests *HsPINK1* to be an active ubiquitin kinase even without Ser228 phosphorylation, which contradicts biochemical analysis. Since AlphaFold2 does not yet predict the impact of post-translational modifications, we interpret the prediction such that it is possible and even likely, that *HsPINK1* can adopt a ubiquitin-phosphorylation-competent conformation consistent with our previous *PhPINK1–Ub* TVLN complex structure¹⁰. **c, d**, We next used AlphaFold2 to predict a dimer of *HsPINK1*. **c**, Strikingly, AlphaFold2 predicts a symmetric dimer with a dimer interface identical to the one shown for *PhPINK1* (compare with **d** and Fig. 3). In fact, we have already validated this arrangement of *HsPINK1* molecules via our Cys166 cross-linking experiments in Fig. 3g. However, in the predicted dimer of *HsPINK1*, the α C helix is kinked and Ser228 does not contact the second molecule. This is different from our conclusions in Extended Data Fig. 8, but again may be a result of not incorporating the effect of Ser228 phosphorylation. We therefore anticipate that unphosphorylated *HsPINK1* can also adopt a conformation with an extended α C helix that places Ser228 into the active site of the dimeric molecule to facilitate autophosphorylation prior to forming the depicted conformation. AlphaFold2 predictions hence support the notion that the activation model proposed in Fig. 5g applies to *HsPINK1*.

Supplementary information

Supplementary Figs. 1 and 2

Supplementary Fig. 1: uncropped gels and blots. Supplementary Fig. 2: repeats of Fig. 5e, used for quantification in Fig. 5f.

Reporting Summary

Supplementary Tables 1 and 2

Crystallographic and cryo-EM data.

Peer Review File

Supplementary Data

Source code: MATLAB script to reproduce Kolmogorov–Smirnov test in Extended Data Fig. 9g.

Supplementary Video 1

YFP–Parkin translocation in *PINK1*^{−/−} HeLa cells expressing *HsPINK1* variants. Live cell imaging of YFP–Parkin (cyan) translocation to the mitochondria (magenta, TOM20–Halo) after OA treatment in *PINK1*^{−/−} HeLa cells stably expressing the indicated *HsPINK1* variants. KD, kinase dead. Compare with Extended Data Fig. 9. Images were taken every minute for 60 min using a Lattice Light Sheet 7 (Zeiss) microscope. Maximum intensity projections are shown for four different timepoints. Images are representative of three independent experiments. Scale bar, 10 μm. Frame rate, approximately 3 fps.

Source data

Source Data Fig. 5

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 9

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

Gan, Z.Y., Callegari, S., Cobbold, S.A. *et al.* Activation mechanism of PINK1. *Nature* **602**, 328–335 (2022). <https://doi.org/10.1038/s41586-021-04340-2>

- Received: 22 August 2021
- Accepted: 13 December 2021
- Published: 21 December 2021
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04340-2>

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

PINK spots: diseased mitochondria prepare for mitophagy

- Sara Osman

News & Views 10 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04340-2>

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

- Article
- Open Access
- [Published: 02 February 2022](#)

Crystallographic snapshots of a B₁₂-dependent radical SAM methyltransferase

- [Cameron D. Fyfe](#)^{1 na1},
- [Noelia Bernardo-García](#)^{1 na1},
- [Laura Fradale](#)^{1 na1},
- [Stéphane Grimaldi](#) ORCID: orcid.org/0000-0002-9559-6112²,
- [Alain Guillot](#)¹,
- [Clémence Brewee](#)¹,
- [Leonard M. G. Chavas](#)^{3,4},
- [Pierre Legrand](#) ORCID: orcid.org/0000-0003-2431-2255³,
- [Alhosna Benjdia](#) ORCID: orcid.org/0000-0002-3544-2541¹ &
- [Olivier Berteau](#) ORCID: orcid.org/0000-0002-3434-5168¹

[Nature](#) volume 602, pages 336–342 (2022)

- 4017 Accesses
- 33 Altmetric
- [Metrics details](#)

Subjects

- [Biocatalysis](#)
- [Biosynthesis](#)
- [Enzyme mechanisms](#)

Abstract

By catalysing the microbial formation of methane, methyl-coenzyme M reductase has a central role in the global levels of this greenhouse gas^{1,2}. The activity of methyl-coenzyme M reductase is profoundly affected by several unique post-translational modifications^{3,4,5,6}, such as a unique C-methylation reaction catalysed by methanogenesis marker protein 10 (Mmp10), a radical S-adenosyl-l-methionine (SAM) enzyme^{7,8}. Here we report the spectroscopic investigation and atomic resolution structure of Mmp10 from *Methanosarcina acetivorans*, a unique B₁₂ (cobalamin)-dependent radical SAM enzyme⁹. The structure of Mmp10 reveals a unique enzyme architecture with four metallic centres and critical structural features involved in the control of catalysis. In addition, the structure of the enzyme–substrate complex offers a glimpse into a B₁₂-dependent radical SAM enzyme in a precatalytic state. By combining electron paramagnetic resonance spectroscopy, structural biology and biochemistry, our study illuminates the mechanism by which the emerging superfamily of B₁₂-dependent radical SAM enzymes catalyse chemically challenging alkylation reactions and identifies distinctive active site rearrangements to provide a structural rationale for the dual use of the SAM cofactor for radical and nucleophilic chemistry.

[Download PDF](#)

Main

Methane production by anaerobic methane-oxidizing archaea is responsible for two-thirds of global methane emissions¹, a large part of which originates from marine sediments¹ and the mammalian microbiome^{2,10}. In this process, methyl-coenzyme M reductase (MCR) has a central role by catalysing the reversible interconversion of 2-methylmercaptoethanesulfonate (CoM) and 7-thioheptanoylthreoninephosphate (CoB) to a CoB–CoM heterodisulfide and methane (Fig. 1a). The structure of MCR³ has revealed several distinct features for this 300-kDa (αβγ)₂ protein complex, such as a unique F₄₃₀ cofactor¹¹ and unusual post-translational modifications⁵, including 5-C-(S)-methylarginine^{4,5}, which tunes the reactivity of its active site^{6,12}. Mmp10, which has been shown to catalyse this key post-translational modification^{7,8}, belongs to an emerging superfamily of B₁₂-dependent radical SAM enzymes^{13,14,15,16,17,18,19,20,21} that encompasses more than 200,000 proteins (<http://radicalssam.org/>)²². These enzymes are involved in the biosynthesis of myriad natural products including bacteriochlorophyll and antibiotics^{9,16,18,23} and catalyse various reactions such as methyl transfer to sp²- and sp³-hybridized carbon atoms^{13,14,18,24}, P-methylation²⁵, ring contraction and cyclization reactions^{26,27}. However, despite the numerous biochemical and spectroscopic studies available in the literature^{13,14,15,16,17,18,19,20,21},

knowledge of these biological catalysts remains limited. Notably, only two structures of B_{12} -dependent radical SAM enzymes have been solved thus far; however, these studies present some limitations precluding a deep understanding of their catalysis^{24,27}. In addition, no structure of a B_{12} -dependent radical SAM enzyme catalysing methyl transfer to an sp^3 -hybridized carbon atom has been reported so far, even though these enzymes are the only known biological catalysts capable of such transformation and this reaction is by far the most widespread in this enzyme family. Interestingly, these latter enzymes also have the remarkable property of being able to make dual use of SAM to initiate radical chemistry and to catalyse nucleophilic displacement, which remains poorly understood.

Fig. 1: MCR and Mmp10 activity with overall structure of Mmp10.

 figure 1



a, The activity of MCR producing CoB–CoM heterodisulfide and methane is enhanced by the post-translational modification of R285 catalysed by the B_{12} -dependent radical SAM enzyme Mmp10. **b**, Overall structure of Mmp10 with bound sodium, [4Fe–4S] cluster, MTA, MeCbl and single iron atom cofactors (Protein Data Bank (PDB) accession 7QBT). Although Mmp10 was crystallized with SAM, only electron density for MTA was observed (Extended Data Fig. 1, Extended Data Table 1). **c**, Magnified view of the [4Fe–4S] cluster coordinated by three cysteine residues and Y115 alongside the modelled MTA molecule not coordinated to the cluster (Extended Data Fig. 1b). **d**, Iron loop with a single iron atom coordinated by four cysteine residues (Extended Data Fig. 1c). Light blue, radical SAM domain; teal, cobalamin-binding domain; purple, iron loop; green, MTA; magenta, MeCbl. The [4Fe–4S] cluster is shown as yellow and orange spheres; the single iron is presented as an orange sphere; and the sodium atom is shown as a violet sphere. Omit maps (blue mesh) of the [4Fe–4S] cluster, its coordinated Y115 and the uncoordinated MTA (**c**) or single iron atom (**d**) contoured at 3σ are depicted.

Mmp10 has a unique architecture

The structure of holo-Mmp10 was solved at an atomic resolution of 1.9 Å, and electron density was obtained for the 411 residues of the protein (Extended Data Table 1). In a unique manner, Mmp10 is composed of two domains and an iron loop (Fig. 1b, Extended Data Fig. 1a). The first domain (residues 1–257) has an unusual radical SAM triosephosphate isomerase (TIM) barrel motif ($\beta_7\alpha_6$). The radical SAM [4Fe–4S] cluster^{28,29,30,31} is coordinated by three cysteine residues (C15, C19 and C22) and, in contrast to all known radical SAM enzymes, by a strictly conserved tyrosine residue (Y115) located between the third β -strand and third α -helix of the TIM barrel (Fig. 1c, Extended Data Figs. 1b, 2). Thus far, FeS cluster coordination by a tyrosine residue has been reported for only the [FeFe]-hydrogenase maturase HydE³² and one nitrogenase³³. With Y115 coordinating the radical SAM cluster, SAM is not able to bind the ‘unique iron’. However, we observed an electron density for the adenine moiety of SAM at its expected location near the top β -barrel sheet (Fig. 1b, Extended Data Fig. 1b) and coordinated by the protein backbone through hydrogen bonds. Notably, despite the atomic resolution, we could not resolve the methionine moiety of SAM, suggesting some flexibility for this cofactor, which was hence modelled as S-methyl-5'-thioadenosine (MTA) in this structure. This flexibility probably occurs because SAM is not coordinated to the [4Fe–4S] cluster and no polar interaction occurs with the glycine-rich motif (GGD), which usually binds the amino group of the methionine moiety of SAM²⁹. Co-crystallization of Mmp10 with S-adenosyl-l-homocysteine (SAH) resulted in electron density consistent with the presence of the full SAH cofactor. However, among the five canonical radical SAM motifs, only a few contacts were observed, with no direct interaction between SAH and the GGD and ribose motifs. In these different structures, the adenine moiety of SAH or MTA is close to the radical SAM cluster (3.8 Å from the nearest ion of the cluster to the adenine moiety), whereas the [4Fe–4S] cluster and cobalamin (vitamin B₁₂) are separated by 12 Å (from the nearest ion of the cluster to the cobalt atom).

The second unique feature of Mmp10 is the presence of a loop coordinating mononuclear iron (Fig. 1d, Extended Data Fig. 1c), which is similar to a rubredoxin iron loop³⁴ although with a distinct orientation. This mononuclear iron is coordinated by a unique and conserved cysteine motif (C35, C38, C45 and C48) within the radical SAM domain (Extended Data Fig. 2). Although the presence of additional FeS clusters is common in non-B₁₂-dependent radical SAM enzymes, no radical SAM enzyme has been shown to contain a mononuclear centre. Furthermore, auxiliary FeS clusters are often located in the C-terminal region and outside the TIM barrel domain, with the notable exception of BioB³⁵, which is built on a complete TIM barrel.

Roles of iron sites in Mmp10

To investigate the properties of the radical SAM cluster and iron loop, we generated an A3 mutant lacking the three cysteine residues from the radical SAM cluster and an A4 mutant lacking the four cysteine residues from the iron loop, along with several mutants with individual alanine substitutions. The activity of these mutants was compared to that of wild-type protein using a synthetic peptide substrate ($[M + H]^+$: 1,496.77) mimicking MCR⁸ and containing R285 (numbered as in MCR), which is the target of the methylation reaction. Mmp10 efficiently transferred a methyl group to R285 in the presence of SAM and Ti(III) citrate (Extended Data Fig. 3a–c), as shown by a mass shift of $\Delta m = +14$ Da ($[M + H]^+$: 1,510.79). One molecule each of 5'-deoxyadenosine (5'-dA) and SAH were produced per methylation reaction (Extended Data Fig. 3b), whereas no SAM cleavage was noted in the absence of peptide, regardless of the reductant used. As expected, the A3 mutant was inactive and unable to cleave SAM (Extended Data Fig. 3d). Mutants of the iron loop (A4 and C38A mutants) produced small amounts of 5'-dA but were unable to transfer a methyl group to the substrate (Extended Data Fig. 3e,f). Finally, when we abrogated the coordination of Y115 to the [4Fe–4S] cluster (Y115A mutant), the cleavage activity towards SAM was severely impaired and the methyltransferase activity was abolished (Extended Data Fig. 3g). Substitution of Y115 with phenylalanine only marginally restored enzyme activity (<1% of wild-type activity), which demonstrates the critical involvement of the hydroxyl group of Y115 in polar interactions following substrate binding (Extended Data Fig. 3h).

Electron paramagnetic resonance (EPR) analysis of Mmp10 revealed unique spectroscopic signatures (Extended Data Fig. 3i). First, oxidized Mmp10 exhibited a strong high-spin signal ($S = 5/2$) at $g = 4.30, 4.14$ and 9.4 , which is characteristic of a mononuclear Fe^{3+} ion, and showed a signal for $[3Fe–4S]^+$ at $g = 2.0$. The latter signal corresponds to the oxidized radical SAM cluster, whereas the high-spin $S = 5/2$ signal is unusual and mirrors those reported in oxidized rubredoxin. After FeS reconstitution and reduction, signals at $g = 2.03, 1.93$ and 1.88 were noted, which correspond to the radical SAM $[4Fe–4S]^+$ cluster (Extended Data Fig. 3i). In the low-field region, signals at $g = 5.4$ and 3.1 are characteristic of spin systems of $S = 3/2$ and are fully consistent with a [4Fe–4S] cluster coordinated by three cysteine residues and a non-cysteine ligation³⁶. The reduced A4 mutant exhibited an EPR spectrum similar to that of the wild-type enzyme although with an altered signal of $S = 3/2$, leading to the appearance of a signal at $g = 1.15$ (Extended Data Fig. 3e). By contrast, mutation of the three cysteine residues from the radical SAM motif abrogated signal for the $S = 3/2$ species (Extended Data Fig. 3d). Finally, the addition of SAM to reduced Mmp10 led to a change in the EPR spectrum, with the development of additional signals at $g = 1.89$ and 1.80 . This result is consistent with direct interaction between SAM and the [4Fe–4S] cluster³⁷ (Extended Data Fig. 3i). Collectively, these data support the idea

that both the $S=1/2$ and $S=3/2$ spin systems originate from the radical SAM [4Fe–4S] cluster and are consistent with the existence in solution of free and Y115-bound forms.

A distinct B₁₂-binding domain

The B₁₂-binding domain (158 residues) is formed by four β-strands and seven α-helices (Figs. 1, 2). This domain comprises most of the polar bonds that hold the cobalamin dimethylbenzimidazole (DMB) tail in place. By contrast, a network of interactions from mainly the iron loop (Y23, F24 and Y47) and the penultimate loop of the radical SAM domain (R210, N217, I220, L221 and N223) coordinate the side chains of the tetrapyrrole ring. Owing to the low number of β-strands and α-helices and an absence of the polar α-helix involved in phosphate binding, this domain is only marginally related to a Rossmann fold. Furthermore, none of the canonical B₁₂-binding motifs such as the His-on (DXHXXG) and SXL motifs³⁸ were identified. A molecule of SAM (or SAH) and Y23 are found between the [4Fe–4S] cluster and cobalamin (Fig. 2a, Extended Data Fig. 4a–e). Y23 interacts with the tetrapyrrole C8 side chain and F24 through a π–π interaction (Fig. 2a), and its hydroxyl group is 4.8 Å from the cobalt atom (upper axial coordination), suggesting a role for Y23 in tuning cobalt reactivity. Although several charged residues have been reported to serve as a lower axial ligand in B₁₂-binding enzymes either directly or through water contact, Mmp10 has a hydrophobic residue (L322) in the lower axial position of cobalamin (Fig. 2a). Because it lacks a lone pair, this residue cannot coordinate the cobalt atom. Its role is hence probably to maintain the pentacoordination of the cobalt centre and to prevent water molecules from interacting with the cobalt atom. In support of this conclusion, despite the high resolution of the structure, we observed no water molecules beneath the cobalamin cofactor, which is shielded by a hydrophobic pocket (Extended Data Fig. 4g, h). This novel binding mode is probably responsible for the atypical planar geometry of the tetrapyrrole ring (Extended Data Fig. 4f).

Fig. 2: Binding of vitamin B₁₂ and S-adenosyl ligands by Mmp10.

 **figure 2**

-
- a**, The C8 side chain of MeCbl is shown in interaction with Y23, I220, L221 and N223 within the radical SAM domain, resulting in a planar tetrapyrrole ring. MeCbl has no lower axial ligand because it is pentacoordinated; however, L322, which resides at 3.9 Å from the cobalt atom, is part of a loop of residues forming a hydrophobic environment for the cobalt ion. Y23 appears at 4.8 Å from the cobalt ion.
- b**, Snapshots of S-adenosyl cofactors within distinct Mmp10 structures. The distances between the sulfur atom of SAH and MTA or SAM and the cobalt ion are indicated by dashed lines. Top left, Mmp10 crystallized with SAH in the absence of peptide

substrate (1: Mmp10 SAH structure; PDB 7QBV). Bottom left and top right, Mmp10 crystallized in the absence of substrate with SAM. Only the density of MTA was observed, which is labelled accordingly (2: Mmp10 MTA_1; PDB 7QBT; 3: Mmp10 MTA_2; PDB 7QBU). Bottom right, Mmp10 crystallized with SAM and its peptide substrate (4: Mmp10–SAM–peptide structure; PDB 7QBS). Light blue and purple, radical SAM domain residues; teal, cobalamin domain; green, SAM, MTA and SAH; magenta, MeCbl. The [4Fe–4S] cluster is shown as orange and yellow spheres. Omit maps (blue mesh) of ligands contoured at 3σ are depicted. See Extended Data Table 1 and Extended Data Fig. 4 for additional information.

Motion of SAM in the active site

No notable overall structural change was observed when Mmp10 was co-crystallized with the demethylated SAM product, with a root mean squared deviation (r.m.s.d.) of 0.37 Å over 408 residues; SAM adenine binding remained mostly unaffected. However, the distance between the sulfur atom of SAM (MTA) and the cobalt atom of cobalamin was shortened from 8.9 Å to 5.6 Å in an alternative SAM orientation (Fig. 2b, Extended Data Fig. 4). The methionine moiety is hence free to move and rotate to a distance compatible with direct methyl transfer from SAM to the cobalt atom. Following substrate binding, marked changes occurred, with displacement of Y115 coordination from the [4Fe–4S] cluster by the carboxylate and amino groups of SAM (Fig. 2b). The distance between the sulfur atom of SAM and the cobalt atom increased to 9.4 Å, whereas that between the sulfur of SAM and the [4Fe–4S] cluster shortened to 3.4 Å. These results demonstrate that SAM can adopt various conformations within the active site. Unexpectedly, coordination of the [4Fe–4S] cluster by Y115 enables the enzyme to discriminate between radical and nucleophilic uses of SAM, without requiring two SAM-binding sites.

Peptide binding and recognition

Co-crystallization of Mmp10 with its substrate revealed clear electron density for eight residues, including R285, the target of the modification (Fig. 3a). Following substrate binding, Mmp10 adopted a closed conformation involving displacement of the α 1a-helix by 11.6 Å and the α 1- to α 4-helices of the radical SAM TIM barrel by as much as 3.4 Å (Extended Data Fig. 5a, b). In addition to coordination of the methionine moiety of SAM to the cluster, numerous polar contacts were established involving the ribose and GGD motifs and additional interaction between Mmp10 and cobalamin.

Unexpectedly, the C2, C7 and C18 side chains of cobalamin established multiple polar interactions with the peptide backbone (Extended Data Fig. 5c). At the entrance of the active site, the peptide backbone formed a sharp twist assisted by two conserved proline residues and a complex network of polar interactions between charged amino

acid side chains and the enzyme backbone (D6, Y56, E54 and G87) (Extended Data Fig. 5c). R285 exhibited an extended side chain that protruded into the enzyme active site, and its C δ atom is at the perfect distance (3.7 Å) and orientation with respect to the C5' atom of SAM, for direct hydrogen atom abstraction (Fig. 3b). The 4.2-Å distance between the C δ atom and the methyl group of cobalamin is also perfectly compatible with direct transfer of the methyl group from methylcobalamin (MeCbl) to the C δ atom. Notably, the guanidinium moiety of R285 was coordinated not only by polar interaction with the protein (E378) and water contacts but also by the SAM cofactor itself through the adenine and ribose moieties (Extended Data Fig. 5d,e). Finally, Y115 became coordinated via hydrogen bonding to E378, enabling SAM to interact with the fourth iron atom of the [4Fe–4S] cluster and radical chemistry to take place.

Fig. 3: Structure of Mmp10 in complex with its peptide substrate.

 figure 3

a, Overview of Mmp10 in complex with peptide substrate, shown in orange (PDB code 7QBS). **b**, Close-up of the Mmp10 active site showing SAM in green, peptide substrate in orange and cobalamin in magenta. The distance between the C5' atom of SAM and the C δ atom of the arginine peptide substrate R285 (3.7 Å) and that between R285 and the methyl group of MeCbl (4.1 Å) are indicated by dashed lines. The omit map (blue mesh) of peptide is contoured at 3 σ . **c**, Sequences of the peptides used as potential substrates with the substitution of arginine (R285) with isoleucine (Ile), leucine (Leu), ornithine (Orn), lysine (Lys), citrulline (Cit) or homoarginine (HArg). **d**, UV-visible light analysis of Mmp10 pre-incubated with OHCbl. Green line, OHCbl–Mmp10; blue line, OHCbl–Mmp10 after incubation with Ti(III) citrate; red line,

reduced OHCbl–Mmp10 exposed to air. See Extended Data Figs. 4–6 for additional information.

Enzyme specificity

Mmp10 introduces only a single modification in MCR, which suggests a strict specificity contrary to that exhibited by enzymes installing multiple post-translational modifications in ribosomally synthesized and post-translationally modified proteins^{9,23,39,40,41,42,43}. To investigate the substrate promiscuity of Mmp10, we substituted R285 with hydrophobic residues (isoleucine or leucine) or structural analogues (Fig. 3c). None of these peptides were modified by Mmp10 despite having a C δ atom in the target side chain (Fig. 3c, Extended Data Fig. 6a, Extended Data Table 2), including a citrulline derivative that differed from the wild-type peptide by only one atom. Competition experiments provided additional support that this analogue does not interact with Mmp10 (Extended Data Fig. 6b), which is consistent with the importance of the guanidinium moiety for interaction with E378 (Extended Data Fig. 5). In addition, E378 coordinated with Y115, which supports the idea that substrate binding acts as a switch for cluster availability. Collectively, a complex network of interactions involving water molecules and SAM, along with protein dynamics, controls the strict specificity of this enzyme. Finally, UV–visible light analysis of the hydroxycobalamin (OHCbl)–enzyme complex showed that, following Ti(III) citrate treatment, a Co(I) intermediate is formed (Fig. 3d), providing a route for MeCbl regeneration.

Proposed mechanism for Mmp10 catalysis

Interaction with the substrate, likely assisted by reduction of the [4Fe–4S] cluster, has a major role in displacement of Y115 from the radical SAM [4Fe–4S] cluster, which enables direct coordination of SAM (Fig. 4). After SAM cleavage, the formed 5'-dA radical abstracts the C δ hydrogen atom of R285, which is at a perfect distance for direct interaction with the methyl group of MeCbl, and induces methyl transfer to R285. Then, Y115 reverts to coordination of the radical SAM [4Fe–4S] cluster, which prevents binding of a novel SAM molecule. The Co(II) intermediate generated during catalysis must be further reduced to produce the super-nucleophile Co(I) for reaction with a second molecule of SAM and to regenerate MeCbl. Interestingly, in the absence of a strong reductant, Mmp10 can convert OHCbl into MeCbl, similar to what has been observed for TsrM¹³, albeit with lower efficiency (Extended Data Fig. 6c). Although the iron loop is ideally located and exposed to solvent, it is unlikely to have a redox potential in the range of the base-off Co(I)–Co(II) redox couple⁴⁴, even though the potential of cobalamin can largely be influenced by the protein matrix⁴⁵.

(Extended Data Fig. 7a). At present, we favour the involvement of a ferredoxin in the reduction of Co(II).

Fig. 4: Proposed mechanism for Mmp10.

 figure 4



Nucleophilic and radical catalysis are highlighted in blue and orange, respectively.

A cation modelled as sodium is present and interacts with the residues holding Y115 in place. Its presence appears to be essential for preventing major backbone reorganization during Y115 motion, as mutation of D156, which makes key interactions in the cation-binding site, severely impairs enzyme activity (Extended Data Fig. 7b,c). The presence of a cation is reminiscent of PFL-AE⁴⁶ and QueE⁴⁷. However, in Mmp10, the cation is not located in the active site. Finally, four *cis*

peptide bonds are present in the structure, including a rare non-proline *cis* peptide bond (Extended Data Fig. 8). These bonds are critical for the interface between the radical SAM and cobalamin-binding domains and are necessary for strict control of catalysis.

Discussion

The structure of Mmp10, the first, to our knowledge, B_{12} -dependent radical SAM enzyme catalysing protein post-translational modification, reveals the mechanism of action of these enzymes in transferring methyl groups to sp^3 -hybridized carbon atoms. Structural and spectroscopic analyses showed that Mmp10 contains four metallic centres at interaction distances (12–16 Å) (Extended Data Fig. 8c). We establish that Mmp10 has a C-terminal B_{12} -binding domain, although residues from the whole protein are involved in binding of the B_{12} cofactor. In addition, this study reports the structure of a B_{12} -dependent radical SAM enzyme in complex with its substrate properly positioned in the active site. This structure provides critical information about the structural and functional diversity of radical SAM enzymes as well as the mechanism of these complex biocatalysts that use both a radical and an S_N2 mechanism. Major and unprecedented active site reorganization occurred following substrate binding. EPR spectroscopy and crystallographic snapshots establish that the radical SAM cluster can be transiently coordinated by a tyrosine residue, which enables the enzyme to perform either radical or nucleophilic chemistry. Recently, the structure of the B_{12} -dependent methyltransferase TsrM^{13,14,24} was solved and shown to contain a [4Fe–4S] cluster coordinated by three cysteines and one glutamate residue²⁴, a coordination encountered in several FeS proteins^{48,49} and proposed to preclude radical catalysis. Our study demonstrates that such coordination in Mmp10 is an intermediate state enabling dual use of the SAM cofactor.

Mmp10 contains a unique iron loop positioned beneath the B_{12} cofactor that is probably involved in shuttling electrons from the cobalt atom. In addition, a hydrophobic pocket prevents water from converting the pentacoordinated MeCbl into its more stabilized hexacoordinated counterpart, which is a strategy conserved in other B_{12} -dependent radical SAM enzymes. These enzymes thus appear to have evolved unique structures and mechanisms to alkylate sp^2 - and sp^3 -hybridized carbon atoms using the twin catalytic power of the cobalamin and SAM cofactors^{8,13,14,16,17,18,19,21,50,51}. In contrast to catalysis by known radical SAM enzymes, catalysis by Mmp10 requires active site reorganization and SAM flexibility within the active site. Although Mmp10 has a unique architecture among known enzymes, the role of such structural rearrangements has probably been underestimated

in radical SAM enzymes, with the current work thereby delineating novel catalytic territories.

Methods

Protein purification

The gene for Mmp10 was commercially synthesized with codon optimization for *Escherichia coli* expression. The *Mmp10* gene was cloned into pET28-a and was transformed into *E. coli* BL21 Star (DE3) cells (Life Technologies) alongside a pRSF plasmid expressing the ISC system. Mutants were generated by gene synthesis or site-directed mutagenesis. Cells were cultured in LB with ampicillin (0.1 mg ml⁻¹) and kanamycin (0.05 mg ml⁻¹). Cultures were incubated at 310 K until the OD₆₀₀ reached 0.7, at which point (NH₄)₂Fe(SO₄)₂ and 0.5 mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) were added to the medium. The cultures were then cooled to 291 K and were incubated for 16 h. Mmp10 was purified by affinity chromatography in 50 mM Tris (pH 8), 400 mM NaCl and 3 mM DTT and was concentrated to ~10 mg ml⁻¹.

Reconstitution of Mmp10 and mutants

Reconstitution and all further sample preparation and experiments were performed in a glovebox in the absence of oxygen. The samples were reconstituted overnight with eightfold molar excess of (NH₄)₂Fe(SO₄)₂ and Na₂S with 3 mM DTT for all experiments unless stated otherwise. Once reconstituted, the samples were buffer-exchanged using PD-10 columns into 50 mM Tris (pH 8), 400 mM NaCl and 1 mM DTT. For crystallography, EPR and UV-visible light analyses, Mmp10 was purified by size exclusion chromatography on a Superdex 200 Increase 10/300GL column using an AKTA Pure system. Samples reconstituted with hydroxocobalamin had tenfold molar excess added to the Mmp10 and were incubated overnight before being passed through a PD-10 column.

Crystallization

The crystallization conditions for holo-native Mmp10 were identified anaerobically at 294 K. Initial crystals appeared after 24 h by using sitting drop diffusion and a 1:1 mixing of protein (10 mg ml⁻¹ with 2 mM SAM and 200 μM MeCbl) and precipitant solutions (100 mM Tris pH 8, 20% polyethylene glycol (PEG) 8000). Holo-native Mmp10 crystals (SAH binding and alternate SAM conformation) were obtained under similar conditions. Holo-native Mmp10 peptide substrate-binding crystals were

optimized using sitting drop vapour diffusion with a 1:1:1 ratio of holo-protein solution with 2 mM substrate peptide (EMLPARRARGPNE) to precipitant solution. The crystals were cryoprotected using 10% PEG 400. All were harvested anaerobically and cryocooled in liquid nitrogen.

Crystallographic structure determination

Diffraction data were collected on the PROXIMA-1 beamline at the synchrotron SOLEIL (Saint-Aubin, France)⁵². A crystal of holo-native Mmp10 with peptide substrate (Mmp10–SAM–peptide structure) belonging to the space group $P6_3$ was detected, and diffraction data were collected to 2.4 Å with phases obtained through multiwavelength anomalous diffraction (MAD). High-energy remote data were collected using an X-ray wavelength of 0.97857 Å and were scaled with a dataset collected at the iron absorption peak at 1.72200 Å. Diffraction images were recorded using an EIGER-X 16M detector, processed with XDS using the XDSME package^{53,54} and corrected for anisotropy using STARANISO⁵⁵. The experimental phasing searched for one FeS cluster site, treated as a super-atom, and one cobalt site using SHARP/AutoSHARP⁵⁶. At this stage, another unexpected separate iron site was found and was included in the phasing. Substructure determination was performed in SHELCXC/D⁵⁷ with heavy atom refinement, phasing and completion performed using SHARP⁵⁸ and density modification using SOLOMON⁵⁹. The model was built using several rounds of automated building with Buccaneer⁶⁰. The final round of model building used ARP/WARP⁶¹, and manual building was performed within Coot⁶² with refinement by Refmac5⁶³ and BUSTER⁶⁴. The final model included the full-length sequence of the protein with one molecule per asymmetric unit. Subsequent data were phased by molecular replacement using PHASER⁶⁵ with this model and with subsequent manual rebuilding and refinement as described above. A holo-native Mmp10 crystal (crystallized with SAM) with the space group $P2_12_12$ diffracted to a resolution of 1.9 Å with four molecules per asymmetric unit (Mmp10 MTA_1 structure). Another crystal (alternate MTA conformation) with the space group $P2_12_12_1$ diffracting to a resolution of 2.3 Å with two molecules per asymmetric unit was also solved (Mmp10 MTA_2 structure). Finally, a further holo-native Mmp10 crystal (crystallized with SAH) with the space group $P2_12_12$ diffracted to a resolution of 2.7 Å and had four molecules per asymmetric unit (Mmp10 SAH structure). Data collection and refinement information can be found in Extended Data Table 1. PyMOL (version 2.0) was used in data analysis and figure generation.

Enzymatic assay with purified enzyme

All reactions were performed anaerobically in the dark. Mmp10 reactions (100–150 µM Mmp10, 3 mM DTT, 200 µM MeCbl, 2 mM SAM, 1 mM peptide substrate, 2 mM Ti(III) citrate) were incubated at 298 K for up to 2 h and were analysed by liquid chromatography–mass spectrometry (LC–MS).

LC–MS analysis

LC–MS analysis was performed using an ultra-high-performance liquid chromatography (UHPLC) instrument (Vanquish Flex, Thermo Scientific) connected by an HESI2 ion source to the MS instrument (Q-Exactive Focus, Thermo Scientific). Samples were diluted 50-fold in buffer A with 2 µl injected onto the column. A reverse-phase column (2.1 mm × 50 mm, 1.7 µm; Eclipse Plus C18, Agilent Technologies) was used for separation. To enhance the retention and resolution of the column, we used heptafluorobutyric acid (HBFA) as an ion-pairing agent with acetonitrile used for buffer B. All compounds eluted between 0% and 50% buffer B during 20 min at a flow rate of 0.3 ml min⁻¹. Buffer A contained 0.2% HBFA in milliQ water; buffer B contained 0.2% HBFA in acetonitrile/MilliQ water at a ratio of 80/20.

EPR spectroscopy

EPR spectra were recorded on a Bruker ElexSys-500 X-band spectrometer equipped with a standard rectangular cavity (ST4102) fitted to an Oxford Instruments liquid helium cryostat (ESR900) and temperature control system. Measurements were conducted at 6 K using a 600-mT or 800-mT field sweep range or at 15 K using a 200-mT field sweep range to focus on the $g = 2.0$ species with a field modulation amplitude of 1 mT at 100 kHz, microwave power of 10 mW and microwave frequency of ~9.48 GHz.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Atomic coordinates and structure factors for the crystal structures reported in this work have been deposited to the Protein Data Bank under the following accession numbers: [7QBS](#) (Mmp10–SAM–peptide structure), [7QBT](#) (Mmp10 MTA_1 structure), [7QBU](#) (Mmp10 MTA_2 structure) and [7QBV](#) (Mmp10–SAH structure).

References

1. Conrad, R. The global methane cycle: recent advances in understanding the microbial processes involved. *Env. Microbiol. Rep.* **1**, 285–292 (2009).
2. Tapiro, I., Snelling, T. J., Strozzi, F. & Wallace, R. J. The ruminal microbiome associated with methane emissions from ruminant livestock. *J. Anim. Sci. Biotech.* **8**, 7 (2017).
3. Ermler, U., Grabarse, W., Shima, S., Goubeaud, M. & Thauer, R. K. Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science* **278**, 1457–1462 (1997).
4. Wagner, T., Wegner, C. E., Kahnt, J., Ermler, U. & Shima, S. Phylogenetic and structural comparisons of the three types of methyl coenzyme M reductase from *Methanococcales* and *Methanobacteriales*. *J. Bacteriol.* **199**, e00197–17 (2017).
5. Kahnt, J. et al. Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea. *FEBS J.* **274**, 4913–4921 (2007).
6. Lyu, Z. et al. Posttranslational methylation of arginine in methyl coenzyme M reductase has a profound impact on both methanogenesis and growth of *Methanococcus maripaludis*. *J. Bacteriol.* **202**, e00654–19 (2020).
7. Deobald, D., Adrian, L., Schone, C., Rother, M. & Layer, G. Identification of a unique radical SAM methyltransferase required for the *sp³*-C-methylation of an arginine residue of methyl-coenzyme M reductase. *Sci. Rep.* **8**, 7404 (2018).
8. Radle, M. I., Miller, D. V., Laremore, T. N. & Booker, S. J. Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical S-adenosylmethionine methylase that unexpectedly requires cobalamin. *J. Biol. Chem.* **294**, 11712–11725 (2019).
9. Benjdia, A., Balty, C. & Berteau, O. Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Front. Chem.* **5**, 87 (2017).
10. Benjdia, A. & Berteau, O. Sulfatases and radical SAM enzymes: emerging themes in glycosaminoglycan metabolism and the human microbiota. *Biochem. Soc. Trans.* **44**, 109–115 (2016).
11. Wongnate, T. et al. The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science* **352**, 953–958 (2016).

12. Nayak, D. D. et al. Functional interactions between posttranslationally modified amino acids of methyl-coenzyme M reductase in *Methanosarcina acetivorans*. *PLoS Biol.* **18**, e3000507 (2020).
13. Pierre, S. et al. Thiomodulon tryptophan methyltransferase expands the chemistry of radical SAM enzymes. *Nat. Chem. Biol.* **8**, 957–959 (2012).
14. Benjdia, A. et al. The thiostrepton A tryptophan methyltransferase TsrM catalyses a cob(II)alamin-dependent methyl transfer reaction. *Nat. Commun.* **6**, 8377 (2015).
15. Kim, H. J. et al. GenK-catalyzed C-6' methylation in the biosynthesis of gentamicin: isolation and characterization of a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **135**, 8093–8096 (2013).
16. Wang, Y. & Begley, T. P. Mechanistic studies on CysS—a vitamin B₁₂-dependent radical SAM methyltransferase involved in the biosynthesis of the *tert*-butyl group of cystobactamid. *J. Am. Chem. Soc.* **142**, 9944–9954 (2020).
17. Wang, Y., Schnell, B., Baumann, S., Muller, R. & Begley, T. P. Biosynthesis of branched alkoxy groups: iterative methyl group alkylation by a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **139**, 1742–1745 (2017).
18. Marous, D. R. et al. Consecutive radical S-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis. *Proc. Natl Acad. Sci. USA* **112**, 10354–10358 (2015).
19. McLaughlin, M. I., Pallitsch, K., Wallner, G., van der Donk, W. A. & Hammerschmidt, F. Overall retention of methyl stereochemistry during B₁₂-dependent radical SAM methyl transfer in fosfomycin biosynthesis. *Biochemistry* **60**, 1587–1596, (2021).
20. Zhong, A., Lee, Y. H., Liu, Y. N. & Liu, H. W. Biosynthesis of oxetanocin-A includes a B₁₂-dependent radical SAM enzyme that can catalyze both oxidative ring contraction and the demethylation of SAM. *Biochemistry* **60**, 537–546 (2021).
21. Parent, A. et al. The B₁₂-radical SAM enzyme PoyC catalyzes valine Cβ-methylation during polytheonamide biosynthesis. *J. Am. Chem. Soc.* **138**, 15515–15518 (2016).
22. Holliday, G. L. et al. Atlas of the radical SAM superfamily: divergent evolution of function using a “plug and play” domain. *Methods Enzymol.* **606**, 1–71 (2018).

23. Benjdia, A. & Berteau, O. Radical SAM enzymes and ribosomally-synthesized and post-translationally modified peptides: a growing importance in the microbiomes. *Front. Chem.* **9**, 678068 (2021).
24. Knox, H. L. et al. Structural basis for non-radical catalysis by TsrM, a radical SAM methylase. *Nat. Chem. Biol.* **17**, 485–491 (2021).
25. Werner, W. J. et al. In vitro phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry* **50**, 8986–8988 (2011).
26. Yang, Z. M. & Bauer, C. E. *Rhodobacter capsulatus* genes involved in early steps of the bacteriochlorophyll biosynthetic pathway. *J. Bacteriol.* **172**, 5001–5010 (1990).
27. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L. & Liu, H. W. A B₁₂-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature* **544**, 322–326 (2017).
28. Berteau, O. & Benjdia, A. DNA repair by the radical SAM enzyme spore photoproduct lyase: from biochemistry to structural investigations. *Photochem. Photobiol.* **93**, 67–77 (2017).
29. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
30. Frey, P. A., Hegeman, A. D. & Ruzicka, F. J. The radical SAM superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88 (2008).
31. Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends Biochem. Sci.* **29**, 243–249, (2004).
32. Rohac, R. et al. Carbon–sulfur bond-forming reaction catalysed by the radical SAM enzyme HydE. *Nat. Chem.* **8**, 491–500 (2016).
33. Owens, C. P., Katz, F. E., Carter, C. H., Oswald, V. F. & Tezcan, F. A. Tyrosine-coordinated P-cluster in *G. diazotrophicus* nitrogenase: evidence for the importance of O-based ligands in conformationally gated electron transfer. *J. Am. Chem. Soc.* **138**, 10124–10127 (2016).
34. Dauter, Z., Wilson, K. S., Sieker, L. C., Moulis, J. M. & Meyer, J. Zinc- and iron-rubredoxins from *Clostridium pasteurianum* at atomic resolution: a high-precision model of a ZnS4 coordination unit in a protein. *Proc. Natl Acad. Sci. USA* **93**, 8836–8840 (1996).

35. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science* **303**, 76–79 (2004).
36. Lanciano, P. et al. New method for the spin quantitation of [4Fe–4S]⁺ clusters with $S = 3/2$. Application to the FS0 center of the NarGHI nitrate reductase from *Escherichia coli*. *J. Phys. Chem. B* **111**, 13632–13637 (2007).
37. Liu, A. & Graslund, A. Electron paramagnetic resonance evidence for a novel interconversion of [3Fe–4S]⁺ and [4Fe–4S]⁺ clusters with endogenous iron and sulfide in anaerobic ribonucleotide reductase activase in vitro. *J. Biol. Chem.* **275**, 12367–12373 (2000).
38. Drennan, C. L., Matthews, R. G. & Ludwig, M. L. Cobalamin-dependent methionine synthase: the structure of a methylcobalamin-binding fragment and implications for other B₁₂-dependent enzymes. *Curr. Opin. Struct. Biol.* **4**, 919–929 (1994).
39. Parent, A. et al. Mechanistic investigations of PoyD, a radical *S*-adenosyl-l-methionine enzyme catalyzing iterative and directional epimerizations in polytheonamide A biosynthesis. *J. Am. Chem. Soc.* **140**, 2469–2477 (2018).
40. Tang, W., Jimenez-Oses, G., Houk, K. N. & van der Donk, W. A. Substrate control in stereoselective lanthionine biosynthesis. *Nat. Chem.* **7**, 57–64 (2015).
41. Mahanta, N., Hudson, G. A. & Mitchell, D. A. Radical SAM enzymes involved in RiPP biosynthesis. *Biochemistry*, <https://doi.org/10.1021/acs.biochem.7b00771> (2017).
42. Benjdia, A., Guillot, A., Ruffié, P., Leprince, J. & Berteau, O. Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in *Bacillus subtilis*. *Nat. Chem.* **9**, 698–707 (2017).
43. Balty, C. et al. Biosynthesis of the sactipeptide ruminococcin C by the human microbiome: mechanistic insights into thioether bond formation by radical SAM enzymes. *J. Biol. Chem.* **295**, 16665–16677 (2020).
44. Banerjee, R. V., Harder, S. R., Ragsdale, S. W. & Matthews, R. G. Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron paramagnetic resonance spectroelectrochemical study. *Biochemistry* **29**, 1129–1135 (1990).

45. Schumacher, W., Holliger, C., Zehnder, A. J. & Hagen, W. R. Redox chemistry of cobalamin and iron-sulfur cofactors in the tetrachloroethene reductase of *Dehalobacter restrictus*. *FEBS Lett.* **409**, 421–425 (1997).
46. Shisler, K. A. et al. Monovalent cation activation of the radical SAM enzyme pyruvate formate-lyase activating enzyme. *J. Am. Chem. Soc.* **139**, 11803–11813 (2017).
47. Dowling, D. P. et al. Radical SAM enzyme QueE defines a new minimal core fold and metal-dependent mechanism. *Nat. Chem. Biol.* **10**, 106–112 (2014).
48. Lee, M. et al. Biosynthesis of isoprenoids: crystal structure of the [4Fe–4S] cluster protein IspG. *J. Mol. Biol.* **404**, 600–610 (2010).
49. Demmer, J. K. et al. Insights into flavin-based electron bifurcation via the NADH-dependent reduced ferredoxin:NADP oxidoreductase structure. *J. Biol. Chem.* **290**, 21985–21995 (2015).
50. Wang, B. et al. Stereochemical and mechanistic investigation of the reaction catalyzed by Fom3 from *Streptomyces fradiae*, a cobalamin-dependent radical S-adenosylmethionine methyltransferase. *Biochemistry* **57**, 4972–4984 (2018).
51. Kim, H. J., Liu, Y. N., McCarty, R. M. & Liu, H. W. Reaction catalyzed by GenK, a cobalamin-dependent radical S-adenosyl-l-methionine methyltransferase in the biosynthetic pathway of gentamicin, proceeds with retention of configuration. *J. Am. Chem. Soc.* **139**, 16084–16087 (2017).
52. Chavas, L. M. G. et al. PROXIMA-1 beamline for macromolecular crystallography measurements at Synchrotron SOLEIL. *J. Synchrotron Radiat.* **28**, 970–976 (2021).
53. Legrand, P. XDSME: XDS made easier. *GitHub* <https://doi.org/10.5281/zenodo.837885> (2017).
54. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
55. Tickle, I. J. et al. STARANISO (Global Phasing, 2018).
56. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
57. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).

58. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
59. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
60. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
61. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
63. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
64. Bricogne, G. et al. BUSTER version X.Y.Z. (Global Phasing, 2017) (2017).
65. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
66. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

Acknowledgements

This work was supported by the European Research Council (ERC consolidator grant 617053), ANR (ANR-17-CE11-0014) and the CHARMMAT Laboratory of Excellence (ANR-11-LABX0039). We express appreciation for use of the EPR facilities available at the French EPR network (IR CNRS3443) and the Aix-Marseille University EPR centre. Use of the beamline PROXIMA-1 at the synchrotron SOLEIL (Saint-Aubin, France) is also acknowledged.

Author information

Author notes

1. These authors contributed equally: Cameron D. Fyfe, Noelia Bernardo-García, Laura Fradale

Affiliations

1. Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, ChemSyBio, Jouy-en-Josas, France

Cameron D. Fyfe, Noelia Bernardo-García, Laura Fradale, Alain Guillot, Clémence Brewee, Alhosna Benjdia & Olivier Berteau

2. Aix Marseille Univ, CNRS, BIP UMR7281, Marseille, France

Stéphane Grimaldi

3. Synchrotron SOLEIL, HelioBio group, L'Orme des Merisiers, Gif sur-Yvette, France

Leonard M. G. Chavas & Pierre Legrand

4. Nagoya University, Nagoya, Japan

Leonard M. G. Chavas

Contributions

A.B. and O.B. designed the research and the experimental strategy. L.F., C.B. and C.D.F. cloned and purified the proteins. L.F. and C.D.F. performed biochemical assays. A.G. analysed reactions by MS. L.F., C.D.F., A.B. and O.B. prepared samples for EPR spectroscopy. S.G. collected and analysed the EPR data. C.D.F., C.B., N.B.G., P.L. and L.M.G.C. crystallized the proteins and collected the crystallographic data. C.D.F., N.B.G., L.F., A.G., L.M.G.C., P.L., A.B. and O.B. analysed and interpreted the crystallographic and biochemical data. C.D.F., A.B. and O.B. wrote the manuscript.

Corresponding authors

Correspondence to [Alhosna Benjdia](#) or [Olivier Berteau](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Qi Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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 Topology diagram of Mmp10 depicting domains \(a\) and 2Fo-Fc and Fo-Fc maps of \(b\) SAM modeled as MTA, \[4Fe-4S\] cluster, Tyrosine 115 and \(c\) the single Fe ion in Mmp10 structure \(Mmp10 MTA_1 structure\).**](#)

Three yellow circles surrounding the [4Fe-4S] cluster represent cysteine residues that coordinate the radical SAM cluster. Purple circle represents Tyrosine 115 that coordinates the fourth iron of the [4Fe-4S] cluster. Four yellow circles within the Fe-loop represent cysteine residues that coordinate a single iron atom shown as an orange circle. Red dot represents Leucine 322 that provides a hydrophobic pocket at the lower axial position of the cobalamin cofactor. Purple sticks represent cobalamin. Pink dot depicts Glutamate 378 involved in substrate binding. Blue dot shows the sodium ion. The 2mFo-DFc map is colored in blue and contoured at 1.0 σ . The mFo-DFc map is colored green (3.0 σ) and red (-3.0 σ).

[**Extended Data Fig. 2 Sequence alignment of Mmp10 homologs.**](#)

258 sequences from the megacluster 1-7 (radicalssam.org), were aligned, and strictly conserved residues are highlighted in colors. For clarity, only eight representative sequences with various degree of similarity from 100 to 37% (indicated in the right column) are shown. The three cysteine residues coordinating the radical SAM cluster, the four cysteine residues coordinating the Fe-loop and Y115 are strictly conserved. Similarly, the residues involved in the B₁₂-binding motif are highly conserved, notably the key L322.

Extended Data Fig. 3 EPR analysis and activity of Mmp10 and mutants.

a, LC-MS analysis of the reaction catalyzed by Mmp10. SAM [M]⁺= 399.14, 5'-dA [M+H]⁺= 252.11 and SAH [M+H]⁺= 385.13 are indicated. The peptide substrate [M+H]⁺= 1496.77 is converted into a methylated species [M+H]⁺= 1510.79 (see panel c for full assignment). **b**, Time course analysis of the reaction catalyzed by Mmp10. **c**, LC-MS/MS analysis of the 13-mer peptide before (upper panel) and after (lower panel) incubation with Mmp10. See Extended Data Table 2 for full assignment. **d**, Schematic representation of the A3-mutant, biochemical characterization and EPR analysis. Right upper panel: LC-MS analysis of the reaction before (blue trace) and after (red trace) 1 hr incubation. Right lower panel: Time course analysis of the reaction. Left lower panel: CW EPR analysis. The weak signal measured at g ~ 2.0 represents adventitious FeS cluster. **e**, Schematic representation, EPR analysis and activity assay of the A4-mutant lacking C35, C38, C45 and C48 involved in the Fe-loop coordination. Upper right panel: CW EPR analysis. Lower panel: LC-MS analysis of the reaction catalyzed by the A4-mutant before (blue trace) and after (red trace) 1 hr incubation. SAM [M]⁺= 399.14, 5'-dA [M+H]⁺= 252.11 and SAH [M+H]⁺= 385.13 are indicated. **f**, Schematic representation of the C38A-mutant and biochemical characterization. Middle panels: LC-MS analysis of the reaction before (blue trace) and after (red trace) 60 min incubation. SAM [M]⁺= 399.14, 5'-dA [M+H]⁺= 252.11 and SAH [M+H]⁺= 385.13 are indicated. Right panel: Time course analysis of the reaction. **g**, Schematic representation of the Y115A-mutant and biochemical characterization. Right and middle panels: LC-MS analysis of the reaction before (blue trace) and after (red trace) 1 hr incubation. SAM [M]⁺= 399.14, 5'-dA [M+H]⁺= 252.11 and SAH [M+H]⁺= 385.13 are indicated. **h**, Schematic representation of the Y115F-mutant and biochemical characterization. Activity of the Y115F mutant in the absence (Left panel) or the presence (middle panel) of the peptide substrate. Right panel: LC-MS analysis of the reaction after 1 hr. All reactions were performed under anaerobic and reducing conditions with reconstituted enzymes and analyzed by LC-MS. Methylated peptide [M+H]⁺= 1510.79 (●, ○), 5'-dA [M+H]⁺= 252.11 (■, □) and SAH [M+H]⁺= 385.13 (▲, Δ) are indicated. Experiments were performed in duplicate. **i**, Schematic representation and EPR analysis of Mmp10. Ox: Oxidized enzyme as purified, Red: Reduced enzyme after anaerobic reconstitution and incubation with sodium dithionite. SAM: Reduced enzyme after incubation with SAM. g-values are indicated on the panels. In all panels, the radical SAM cluster ligated by Tyr-115, Cys-15, Cys-19 & Cys-22, the iron-loop coordinated by Cys-35, Cys-38, Cys-45 and Cys-48 and the cobalamin cofactor are indicated.

Extended Data Fig. 4 Methylcobalamin, MTA, SAH and SAM in Mmp10 structures.

2Fo-Fc and Fo-Fc maps of Mmp10 crystallized with SAH in the absence of peptide substrate (**a: Mmp10 SAH structure**). Mmp10 crystallized in the absence of substrate with SAM. Only the density of MTA was observed and labeled accordingly (**b & c**: Mmp10 MTA_1 and d: Mmp10 MTA_2 structures). Mmp10 crystallized with SAM and its peptide substrate (**e**: Mmp10 SAM peptide structure). The 2mFo-DFc map is colored blue and contoured at 1.0 σ . The mFo-DFc map is colored green (3.0 σ) and red (-3.0 σ). See Extended Data Table 1 & Fig. 2 for additional information. **f**, Cobalamin bound within Mmp10 (left; PDB 7QBT) and TsrM (right; PDB 6WTF). Cobalamin from TsrM shown in green and methylcobalamin from Mmp10 shown in magenta. The corrin ring and the sidechain C8 exhibit distinct conformations. **g & h**, Hydrophobic pocket of Mmp10 binding methylcobalamin (Mmp10 MTA_1 structure). The pocket directly beneath the cobalt of methylcobalamin has no water molecules and no charged sidechains, resulting in a hydrophobic environment. The distance between L322 and the cobalt ion is of 4 Å. Surface charge generated using APBS⁶⁶ surface charge within PyMOL (Schrödinger).

Extended Data Fig. 5 Mmp10 interaction with its substrate.

a, Structure superimposition of substrate-free with SAH and peptide-bound with SAM structures of Mmp10. **b**, Enlargement showing the major structural movements including the α 1a helix and the α 1 to α 4 helices of the TIM barrel. Mmp10 with peptide bound is shown in grey and without peptide in teal. Structures were aligned according to the B12 binding domain. Alignment of the peptide-bound and substrate-free SAH structures using all domains (403 residues) has a r.m.s.d. of 0.963 Å. **c**, Polar interactions between Mmp10, its cofactors (methylcobalamin & SAM) and peptide substrate. Omit map of peptide substrate contoured at 3 σ level (in blue). SAM is shown in green, methylcobalamin is depicted in magenta, bound peptide is shown in orange, polar contacts are shown as grey dotted lines, water is shown as a red sphere. **d**, Omit map of the peptide substrate in Mmp10 active site contoured at 3 σ level (in light blue). **e**, The 2mFo-DFc map of the peptide substrate in Mmp10 active site is colored blue and contoured at 1.0 σ . The mFo-DFc map is colored green (3.0 σ) and red (-3.0 σ). SAM shown in green, methylcobalamin depicted in magenta, peptide substrate in orange.

Extended Data Fig. 6 Activity of Mmp10 on peptide substrate.

a, Influence of the critical R285 residue. Reactions were analyzed by LC-MS. Only for the wild-type peptide a methyl transfer reaction was observed (mass shift of $\Delta m = +14$ Da). See Extended Data Table 2 for peptide masses and Extended methods for reaction conditions. **b**, Activity of Mmp10 incubated with the wild-type peptide alone (left panel) or the wild-type and citrulline-containing peptides (right panel). Reactions were analyzed by LC-MS. Methylated peptide $[M+H]^+ = 1510.79$ (circles), 5'-dA $[M+H]^+ =$

252.11 (squares) and SAH $[M+H]^{+}=385.13$ (triangles). Experiments were performed in duplicate. **c**, Formation of MeCbl by Mmp10 in the absence or the presence of Ti(III)citrate. Mmp10 was incubated under anaerobic and reducing conditions with SAM (1 mM), OHcbl (100 μ M) and DTT (6 mM). Diamonds: Reaction with 1mM Ti(III)citrate. Squares: Reaction without Ti(III)citrate. Circles: Control reaction without enzyme and Ti(III)citrate. Experiments were performed in duplicate. MeCbl was detected by LC-MS analysis $[M+H]^{2+}$: 672.80 and comparison with its retention time with authentic standard.

Extended Data Fig. 7 Fe loop and cation site in Mmp10.

a, Fe loop in Mmp10 (Mmp10 MTA_1 structure). Cobalamin colored magenta, residues from Fe loop colored purple and residue from cobalamin binding domain colored teal. **b**, Cation site in Mmp10 (Mmp10 MTA_1 structure). Na ion (purple sphere) modelled in cation site with an average distance of 2.4 \AA in octahedral conformation (grey dotted lines) from coordinating sidechains and water. Distance between Na ion and 4Fe4S cluster is 11 \AA (yellow dotted line). Omit map of Na ion contoured at 3 σ level. 4Fe4S cluster shown as orange and yellow spheres next to SAM molecule depicted in green. **c**, Activity of the D156A mutant (upper panel) compared to the WT enzyme (lower panel). Reactions were analyzed by LC-MS. Methylated peptide $[M+H]^{+}=1510.79$ (circles), 5'-dA $[M+H]^{+}=252.11$ (squares) and SAH $[M+H]^{+}=385.13$ (triangles). Experiments were performed in duplicate.

Extended Data Fig. 8 Cis-peptide bond and metallic centers in Mmp10.

a, Cis-peptide bonds at the interface between the radical SAM and the cobalamin binding domains in the substrate free (Mmp10 MTA_1 structure) and **b**, peptide-bound Mmp10 structures. Peptide substrate shown in orange with the radical SAM domain coloured in light blue and cobalamin binding domain in teal. Polar interactions shown as yellow staggered lines. Cis-peptide bond between T257 and P258 are coordinating a water molecule which is also involved in substrate binding. The rare non-proline cis-peptide bond found between L259 and E260 is held in place through interactions between the sidechain nitrogens of R243 and R298 and the side chain oxygens of E260. **c**, Metallic centres in Mmp10. Mmp10 SAM peptide structure. Peptide substrate shown in orange.

Extended Data Table 1 Data collection and refinement statistics

Extended Data Table 2 Mass spectrometry data

Supplementary information

Reporting Summary

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

Fyfe, C.D., Bernardo-García, N., Fradale, L. *et al.* Crystallographic snapshots of a B₁₂-dependent radical SAM methyltransferase. *Nature* **602**, 336–342 (2022).

<https://doi.org/10.1038/s41586-021-04355-9>

- Received: 25 June 2021
- Accepted: 15 December 2021
- Published: 02 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04355-9>

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/s41586-021-04355-9>

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

- Article
- [Published: 02 February 2022](#)

Structure of a B₁₂-dependent radical SAM enzyme in carbapenem biosynthesis

- [Hayley L. Knox](#) ORCID: [orcid.org/0000-0003-1467-309X¹](https://orcid.org/0000-0003-1467-309X)^{nal},
- [Erica K. Sinner](#) ORCID: [orcid.org/0000-0003-2417-9319²](https://orcid.org/0000-0003-2417-9319)^{nal},
- [Craig A. Townsend](#) ORCID: [orcid.org/0000-0003-2795-7585²](https://orcid.org/0000-0003-2795-7585),
- [Amie K. Boal](#) ORCID: [orcid.org/0000-0002-1234-8472^{1,3}](https://orcid.org/0000-0002-1234-8472) &
- [Squire J. Booker](#) ORCID: [orcid.org/0000-0002-7211-5937^{1,3,4}](https://orcid.org/0000-0002-7211-5937)

Nature volume 602, pages 343–348 (2022)

- 7160 Accesses
- 99 Altmetric
- [Metrics details](#)

Subjects

- [Antibiotics](#)
- [Metalloproteins](#)
- [X-ray crystallography](#)

Abstract

Carbapenems are antibiotics of last resort in the clinic. Owing to their potency and broad-spectrum activity, they are an important part of the antibiotic arsenal. The vital role of carbapenems is exemplified by the approval acquired by Merck from the US Food and Drug Administration (FDA) for the use of an imipenem combination therapy to treat the increased levels of hospital-acquired and ventilator-associated bacterial pneumonia that have occurred during the COVID-19 pandemic¹. The C6 hydroxyethyl side chain distinguishes the clinically used carbapenems from the other classes of β-

lactam antibiotics and is responsible for their low susceptibility to inactivation by occluding water from the β -lactamase active site². The construction of the C6 hydroxyethyl side chain is mediated by cobalamin- or B₁₂-dependent radical S-adenosylmethionine (SAM) enzymes³. These radical SAM methylases (RSMTs) assemble the alkyl backbone by sequential methylation reactions, and thereby underlie the therapeutic usefulness of clinically used carbapenems. Here we present X-ray crystal structures of TokK, a B₁₂-dependent RSMT that catalyses three-sequential methylations during the biosynthesis of asparenomycin A. These structures, which contain the two metallocofactors of the enzyme and were determined in the presence and absence of a carbapenam substrate, provide a visualization of a B₁₂-dependent RSMT that uses the radical mechanism that is shared by most of these enzymes. The structures provide insight into the stereochemistry of initial C6 methylation and suggest that substrate positioning governs the rate of each methylation event.

[Download PDF](#)

Main

During the biosynthesis of all known complex carbapenem natural products, the assembly of the C6 alkyl side chain (Fig. 1a) is accomplished by a cobalamin (Cbl or B₁₂)-dependent radical SAM enzyme^{3,4,5}. These catalysts can perform serial methyl transfers with control of stereochemical outcome for each reaction. The Cbl-dependent RSMT ThnK performs two sequential methyl transfers to its (2R)-pantetheinylated carbapenam substrate **1** during the biosynthesis of the paradigm carbapenem antibiotic thienamycin (**2**) (ref. ⁶). An orthologue of ThnK, TokK from *Streptomyces tokunonensis* (ATCC 31569), constructs the C6 isopropyl chain of the carbapenem, asparenomycin A (**3**), by deploying three sequential methylations of **1** (Fig. 1b). This biosynthetic approach allows the producing organism to make a small ‘library’ of alkylated analogues, which may deter the development of resistance in susceptible bacteria. This strategy may also be used in the biosynthesis of cystobactamids, in which a similar Cbl-dependent radical SAM (RS) enzyme, CysS, performs successive methyl transfers⁷ (Extended Data Fig. 1). Notably, despite low sequence identity (around 29%) between CysS and TokK or ThnK, all three proteins are located within the same cluster of a sequence similarity network (SSN) composed of approximately 11,000 Cbl-binding RS enzymes obtained from networks provided by radicalSAM.org (<https://radicalSAM.org/>) and the UCSF structure–function linkage database (SFLD) (Fig. 1c, Supplementary Fig. 1). It is tempting to speculate that this colocalization might be driven by mechanistic similarities.

Fig 1: Cbl-dependent radical-mediated methylations in carbapenem biosynthesis.

 **figure 1**

a, Structure of (*2R*)-pantetheinylated carbapenam precursor substrate **1** modified by ThnK and TokK. Related carbapenem natural products containing C6-alkyl substituents include thienamycin and asparenomycin A. **b**, Proposed mechanism describing the three sequential methylations catalysed by TokK. **c**, An abbreviated

SSN of Cbl-binding RS enzymes, highlighting selected sequence clusters and nodes. The full network (Supplementary Fig. 1) was generated from around 11,000 annotated Cbl-dependent RS enzymes with an alignment score of 65. Each node represents a single sequence or a set of sequences with more than 40% sequence identity. Sequence clusters and nodes are coloured by predicted reaction mechanism (see key in the figure). Nodes containing functionally annotated sequences are indicated in colour. Structurally characterized enzymes are represented by enlarged nodes and labeled in boldface.

Carbapenem C6 alkyl chain construction requires stereoselective formation of carbon–carbon bonds between unactivated sp^3 -hybridized carbons. Cbl-dependent RSMTs are the only known biological catalysts capable of such transformations. The Cbl-containing subfamily, depicted as an SSN in Fig. 1c, is also one of the largest in the RS superfamily, a diverse group that functions in the biosynthesis of chlorophyll, lipids and natural products with antiproliferative biological activity^{8,9,10,11}. Although most Cbl-dependent RS enzymes have unknown functions, those that have been characterized are generally—but not exclusively—methylases that act on carbon or phosphorus centres by using methylcobalamin (MeCbl) as an intermediate methyl donor. All RS enzymes, with a single known exception^{12,13,14}, reductively cleave SAM to generate methionine (Met) and a 5'-deoxyadenosyl 5'-radical (5'-dA•) (Fig. 1b). The latter reactive intermediate typically initiates catalysis with a target substrate by abstracting a hydrogen atom. In B₁₂-dependent RSMTs, the substrate radical attacks the methyl group of MeCbl, inducing homolytic cleavage of the cobalt–carbon bond to yield cob(II)alamin and the methylated product (Fig. 1b). After dissociation of the methylated product, Met and 5'-deoxyadenosine (5'-dAH), and rebinding of another molecule of SAM, cob(II)alamin is reduced to cob(I)alamin. Co(I) is a supernucleophile, which acquires a methyl group from SAM to regenerate MeCbl (refs. ^{10,11}) (Fig. 1b).

Two Cbl-dependent RS enzymes have been structurally characterized, TsrM and OxsB (refs. ^{12,15}) (Fig. 1c, Supplementary Fig. 1), which are involved in the biosynthesis of the antibiotics thiostrepton A and oxetanocin A, respectively. Both enzymes are mechanistic outliers among Cbl-dependent RS enzymes and are found in SSN clusters distinct from each other and from TokK (Fig. 1c). OxsB uses Cbl in an unknown manner to catalyse a complex ring contraction of 2'-deoxyadenosine monophosphate (dAMP)¹⁵ (Extended Data Fig. 2a). TsrM methylates an sp^2 -hybridized carbon, C2, of L-tryptophan (Trp) by a polar mechanism¹² (Extended Data Fig. 2b). TsrM is distinctive among all RS enzymes because it does not catalyse the formation of 5'-dA• during catalysis¹². Instead, TsrM uses SAM's carboxylate moiety as an acceptor of the N1 proton of Trp during C2 electrophilic substitution by MeCbl^{12,16}. In addition, the structures of TsrM and OxsB have limitations that prevent full understanding of Cbl-

dependent RS catalysis. The structure of OxsB lacks the dAMP substrate¹⁵. TsrM has been co-crystallized with aza-SAM (a SAM analogue) and Trp, but the Trp substrate is bound in an unproductive conformation, requiring computational docking to understand the structural basis for the reaction outcome.

TokK was crystallized under anoxic conditions in the presence of 5'-dAH and Met, the products of reductive SAM cleavage (Supplementary Fig. 2). Structures of this complex were solved in the absence and presence of substrate **1** to resolutions of 1.79 Å and 1.94 Å, respectively (Extended Data Table 1). TokK shares in common with TsrM and OxsB an N-terminal Cbl-binding domain and a central RS domain containing a [4Fe–4S] cluster (Extended Data Figs. 3, 4). In both TokK structures, Met binds to the unique iron of the [4Fe–4S] cluster, and the position of the 5'-carbon of 5'-dAH suggests that the binding of SAM in TokK is almost identical to that in OxsB, but quite different from that observed in the TsrM from *Kitasatospora setae* (*KsTsrM*) (Extended Data Fig. 4d). A third, C-terminal domain is distinct to TokK^{12,15} (Fig. 2a, Extended Data Fig. 5). Although as-isolated TokK contains MeCbl, hydroxycobalamin (OHCbl) and adenosylcobalamin (AdoCbl), only OHCbl is observed bound to the N-terminal domain in the X-ray crystal structures. This assignment was confirmed by high-resolution mass spectrometry of dissolved crystals (Supplementary Fig. 3). In the structure of the TokK–OHCbl–5'-dAH–Met–substrate complex, which mimics the complex immediately before reaction with substrate, we observed clear $F_o - F_c$ electron density consistent with the shape and size of **1** in one of two monomers in the asymmetric unit (Fig. 2b). In the second monomer, this electron density was also present, but it was of insufficient intensity for substrate modelling (Supplementary Fig. 4). In the chain with substrate bound, the pantetheine tail of **1** occupies a channel that leads from the surface of the protein into the active site (Fig. 2a, Supplementary Fig. 5). This cavity is formed at the interface of all three domains of TokK. The N-terminal Cbl-binding domain and unique C-terminal domain contribute most interactions with the pantetheine unit (Fig. 2a, Extended Data Fig. 6). These include hydrophobic contacts, water-mediated H-bonding interactions and direct polar contacts. For example, Asn515 in the C-terminal domain H-bonds to the terminal -OH of the pantetheine moiety, suggesting that this domain has a key role in substrate recognition. In the N-terminal domain, the Cbl cofactor itself participates in a water-mediated contact to an amide carbonyl of **1**. This network involves one of the Cbl propionamide substituents and Tyr410 of the RS domain. The β-lactam ring of **1** is buried deep within the RS domain and anchored by direct and water-mediated contacts to the C7 carbonyl and C3 carboxylate substituents. This mode of β-lactam interaction resembles that of carbapenem synthase, the enzyme responsible for inversion of stereochemistry at C5 in simple carbapenems¹⁷ (Extended Data Fig. 7). Both enzymes share the use of H-atom abstraction chemistry to selectively target an unactivated C–H bond within the bicyclic β-lactam core, consistent with their conserved substrate-anchoring strategies. The C3 carboxylate of **1** H-bonds to Arg280 in the RS domain

and Tyr652 in the C-terminal domain. These side chains move considerably from their positions in the structure without substrate bound (Fig. [2a](#)), and substitution of Arg280 with Gln results in near complete loss of activity (Fig. [2c](#)), confirming the importance of these side chains in substrate binding.

Fig. 2: TokK binds its carbapenam substrate at the interface of three domains.

 **figure 2**

a, The overall structure of TokK (chain A) is illustrated as a ribbon diagram and coloured by domain. The Cbl-binding Rossmann fold is shown in teal with the Cbl cofactor in stick format and coloured by atom type. The RS domain is shown in light blue. A [4Fe–4S] cluster is shown in orange and yellow spheres. 5'-dAH and Met

coproducts are shown in stick format. The C-terminal domain is shown in pink. Carbapenam substrate (**1**) is shown in light blue sticks, coloured by atom type. **b**, An $F_o - F_c$ omit electron density map is shown for **1** (blue mesh, contoured at 3.0σ). Substrates, cofactors and coproducts are shown in stick format. Distances between reactive groups are given in units of Å. **c**, Product formation for the R280Q TokK variant performed in triplicate (each replicate shown as a symbol). Substrate is shown in black spheres and the first methylated product is shown in pink squares. All activity assays were conducted with substrate **1**, shown in Fig. [1a](#). **d**, Projections of the additional methyl groups added to C6 of **1** and their respective distances (in Å) from the 5' carbon of 5'-dAH and the hydroxyl moiety of OHcbl, the latter of which serves as a surrogate of the active MeCbl cofactor. Spheres labelled Me and Et represent the suggested positions of the newly installed carbon atoms in the mono-methylated and dimethylated products.

The interactions between TokK and substrate **1** position the β -lactam appropriately both for activation of C6 by 5'-dA \cdot and for subsequent methyl addition by the Cbl cofactor¹⁸ (Fig. [2d](#)). C6 of **1** is located directly in front of the 5'-carbon of 5'-dAH, 3.7 Å away, like other RS enzyme–substrate complexes that initiate reactions by H-atom abstraction. The orientation of these two groups in the structure does not reveal whether the pro-*R* or pro-*S* H-atom is removed from C6 of **1** by 5'-dA \cdot , as these substituents project equally above and below the 5'-carbon of 5'-dAH (Fig. [2d](#)). However, the structure does provide insight into the trajectory of methyl addition. C6 of **1** is located 4.2 Å above the axial ligand of Cbl (Fig. [2d](#)) at an angle of around 85° relative to the 5'-carbon of 5'-dAH. This arrangement suggests that the methyl group adds to the bottom face of the β -lactam ring, consistent with the absolute configurations observed in the TokK products and thienamycin^{18,19}. The distance and orientation of reactant functional groups in TokK also compares favourably to other enzymes that catalyse radical-mediated activation and functionalization of a substrate C–H bond, such as iron-dependent hydroxylases in the cytochrome P450 and iron(II)-oxo-glutarate-(Fe-2OG)-dependent superfamilies (Extended Data Fig. [2](#)). These systems orient their reactive groups similarly, but over a slightly shorter distance range^{20,21}. This comparison highlights an important distinction between RSMTs and other radical functionalization enzymes. In P450s and Fe-2OG enzymes, a single reactive entity—a high-valent iron(IV)-oxo or iron(III)-hydroxo group—must both activate substrate and functionalize it. This strategy is inherently limiting because the enzyme can only activate and functionalize substrate from the same side. The Cbl-dependent RS radical functionalization platform is more versatile because the radical activation step is separated from methylation, which allows for more diverse stereochemical outcomes.

Notably, the structure of TokK in complex with the substrate has marked structural similarities to another well-characterized RS methylase that does not rely on MeCbl,

RlmN (ref. 22) (Extended Data Fig. 8). RlmN uses an *S*-methyl cysteinyl (methylCys) residue as an intermediate methyl carrier during the methylation of the sp^2 -hybridized C2 atoms of adenosine 2503 in ribosomal RNA and adenosine 37 in several transfer RNAs (tRNAs). When the structure of RlmN crosslinked to an *Escherichia coli* tRNA^{Glu} substrate is compared to that of substrate-bound TokK, the methylCys residue in the RlmN structure is in a position similar to that of the hydroxyl group of OHCbl in the TokK structure. Moreover, their respective substrates occupy similar positions in the active site (Extended Data Fig. 8). Although the catalytic mechanisms of these two enzymes are distinct, both obey a ping-pong kinetic model, in which one SAM molecule is used to methylate the intermediate methyl carrier, while a second SAM molecule is used to generate a 5'-dA•.

Cbl is multifunctional in TokK, mediating both the polar methylation of Co(I) by SAM and the transfer of a methyl radical to C6 of the substrate. It is bound at the interface of the Cbl and RS domains with its dimethylbenzimidazole base tucked into the Rossmann fold of the N-terminal domain, a conformation termed ‘base-off’ (Fig. 3a, Extended Data Fig. 5). OxsB and TsrM use a similar base-off approach to interact with their Cbl cofactors (Fig. 3a, Extended Data Fig. 4), a binding mode that allows for extensive modulation of the reactivity of the Co(III) ion of MeCbl by the local protein environment²³. In TsrM, this structural feature is essential for the atypical polar methylation of its substrate, Trp, which requires heterolytic cleavage of the Co(III)-carbon bond of MeCbl. The bottom face of the Co ion in TsrM is adjacent to Arg69 but not directly coordinated, which is likely to promote nucleophilic attack of MeCbl by Trp by blocking coordination of a sixth ligand and destabilizing the Co(III)-C bond owing to charge–charge repulsion^{24,25,26} (Fig. 3a). In TokK, a different side chain, Trp76, occupies the lower axial face of the Cbl, residing 3.8 Å from the metal ion (Fig. 3a). Neither Arg69 nor Trp76 lie in a canonical DXHXXG motif exemplified by methionine synthase, wherein the His residue in the motif ligates to the Co, although both are found in the loop following β3 in the Rossmann fold. To investigate the role of this residue, Trp76 was substituted with Phe and Ala. Rates of methylation slightly increased for both site-specific substitutions, and analysis by electron paramagnetic resonance (EPR) spectroscopy suggested that both variants exhibited the same four-coordinate geometry as wild-type TokK (Supplementary Fig. 6). These data suggest that even when the steric bulk of Trp76 is reduced, water does not coordinate the Cbl (Extended Data Fig. 9). To perturb the local environment of the Cbl cofactor further, Trp76 was also substituted with His and Lys. The activity of Trp76His TokK resembles the activities of the Phe and Ala replacements, but the activity of Trp76Lys TokK was reduced by a factor of around 50 for all three methylation steps (Fig. 3b). The substitution tolerance of Trp76 in TokK contrasts with that of Arg69 of TsrM, which when substituted with Lys was unable to transfer a methyl group from MeCbl to substrate¹². Although Trp76 is not widely conserved among other well-characterized

Cbl-dependent RS methylases, it is found in the same sequence context in the CysS primary structure (Extended Data Fig. 1).

Fig. 3: The Cbl- and substrate-binding sites influence overall activity and the relative rates of each TokK methylation step.

 figure 3

a, Comparative analysis of the side chains proximal to the Co ion in two Cbl-binding RS enzymes, TokK and *KsTsrM* (Protein Data Bank (PDB) ID: 6WTF). Selected amino acid side chains are shown in stick format and the Co ion is shown as a pink sphere. **b**, Top, schematic of the three sequential methylations performed by TokK (SPant, pantetheine (Fig. 1a)). Forty-eight-hour time-course experiments performed in triplicate (each replicate represented as a symbol), tracking the product formation of substrate (black spheres), methyl (pink squares), ethyl (purple triangles) and isopropyl (blue diamonds). Bottom, each product was estimated using COPASI (irreversible mass action model using the reaction scheme shown above) and simulated using Virtual Cell (shown as lines) for wild type (WT), W76F, W76A, W76K, W76H, L383F, W215F, E19A/Y20V, W215Y and W215A.

The structure of **1** bound to TokK also rationalizes established differences in rate constants for each of the three methyl transfers catalysed by this enzyme (Fig. 3b). The second methylation to form the ethyl-containing carbapenam product **5** proceeds at least threefold faster than the formation of **4**—a pattern that runs counter to known differences in the reactivity of secondary and primary C–H bonds. Although we do not report a structure containing the singly methylated intermediate **4**, if we presume that **4** remains anchored to Arg280, the methyl group at C6 would be positioned closer to the Cbl axial ligand and potentially at a more optimal angle than the original C6 C–H target. A third methylation to form the isopropyl carbapenem product **6** requires hydrogen atom abstraction from the same carbon, but the newly added ethyl carbon restricts the population of conformers accessible to 5'-dA \bullet . This steric demand could help to explain why the estimated first-order rate constant for the third methylation, k_3 , is slower than the first two methyl transfers, k_1 and k_2 . The buried location of **1**, 5'-dAH, Met and the Cbl cofactor suggests that dissociation of the methylated carbapenam products must occur before dissociation of the SAM cleavage products. This arrangement is consistent with the non-processive kinetic model used to fit the time-course kinetics of each TokK methylation¹⁸. A similar mechanism was proposed for CysS (ref. 7) (Extended Data Fig. 1). Although CysS is only 29% identical both to TokK and to ThnK, all three proteins potentially contain a Trp side chain adjacent both to the Cbl and to the substrate (Extended Data Fig. 1). Substitution of Trp215 with Phe, Ala, or Tyr markedly slows substrate methylation by TokK, which suggests that it could have a role in catalysis (Fig. 3b).

ThnK and TokK share 79.3% sequence identity and act on the same substrate, (2R)-pantetheinylated carbapenam (Supplementary Fig. 7), but ThnK performs two sequential methylations whereas TokK catalyses three^{18,27}. Nearly all residues in proximity to the active site are identical in the two orthologues. However, three non-conserved amino acids near the active site were examined to determine their role in controlling the extent of methylation. Leu383 is positioned deep in the active site and near 5'-dAH (Supplementary Fig. 7). When this residue is substituted with Phe, which

is found at the same position in the primary structure of ThnK (Supplementary Fig. 7), the rate constants for all three methyl transfers are reduced (2.3-, 2.5- and 4.5-fold for k_1 , k_2 and k_3 , respectively) compared to those of wild-type TokK (Fig. 3b). Two adjacent residues at the entrance to the pantetheine-binding tunnel, Glu19 and Tyr20, (Fig. 2a, Supplementary Fig. 7) were replaced with the cognate residues in ThnK to generate an E19A/Y20V double substitution. In this variant, the rate constant for the first methylation is increased 1.4-fold compared to that of the wild type, and the rate constants for the second and third methylations are decreased 1.4- and 3.4-fold, respectively, therefore shifting the kinetic profile closer to the pattern observed with ThnK ($k_1 > k_2$, $k_3 = 0$) (ref. 18) (Fig. 3b).

The structure of TokK solved in the absence of substrate reveals very few differences in overall fold or domain organization relative to the TokK–1 complex (root mean square deviation (RMSD) of 0.53 Å over 603 residues by C α atoms). The substrate-binding channel, located at the interface of the three domains of TokK, remains intact without substrate with only modest alterations in size caused by the aforementioned conformational changes in the side chains of Arg280 and Tyr652 (Fig. 2a, Supplementary Fig. 5). The preformed nature of the substrate-binding tunnel in TokK contrasts with observations from structures of TsrM in the absence and presence of its substrate, in which a loop from the C terminus moves to cap the active site in the presence of Trp¹². Although the C-terminal domain of TokK is considerably larger than that of TsrM, these domains appear to share a common role in substrate interaction (Extended Data Fig. 4).

The overwhelming majority of known Cbl-dependent RSMTs operate by the radical mechanism used by TokK, producing an equivalent of 5'-dAH and SAH for each methylated product molecule. The structure of the TokK active site reveals a scaffold for positioning the cofactors responsible for substrate activation and methyl transfer and is consistent with the non-processive mechanism of sequential methylations observed for the enzyme, which requires the release of each partially alkylated intermediate and both SAM coproducts before reloading the active site for subsequent methylation. Moreover, the structure reveals that there is little active participation in catalysis from other amino acids in the active site, apart from substrate or cofactor binding. This scaffolding approach, which is underscored by the notable absence of conformational changes after substrate binding, may be shared by the only other Cbl-dependent RS enzyme that has, to our knowledge, been structurally characterized in complex with its substrate, TsrM. Although the electrophilic substitution mechanism used by TsrM requires a general base to accept the N1 proton of the indole ring of Trp during catalysis, biochemical studies and models of the functional enzyme–substrate complex suggest that the carboxylate group of a cosubstrate, SAM, functions in this capacity instead of an active-site amino acid. Elucidation of additional structures and mechanisms for Cbl-dependent RS enzymes will reveal how a seemingly common

approach for catalysis may be further elaborated within this large and diverse group of enzymes.

Methods

Overexpression and purification of TokK from *Streptomyces tokunonensis*

TokK (Uniprot ID: A0A6B9HEI0) was produced heterologously in *Escherichia coli* BL21 (DE3) by overexpression from a pET29b:*tokKTev* construct¹⁸. To facilitate production of soluble TokK protein with maximal occupancy of [4Fe–4S] cluster and Cbl cofactors, the strain was transformed with two additional plasmids, pDB1282 and pBAD42-BtuCEDFB^{18,28,29,30}. A 100-ml LB starter culture with 50 µg ml⁻¹ kanamycin (pET29b) containing *tokK*, 50 µg ml⁻¹ spectinomycin (pBAD42-BtuCEDFB) and 100 µg ml⁻¹ ampicillin (pDB1282) was inoculated from a single colony and incubated for 18 h at 37 °C while shaking at 250 rpm. A 12-ml aliquot of the starter culture was used to inoculate a 4-l culture of LB medium supplemented with OHcbl (1.3 µM) and allowed to shake at 180 rpm at 37 °C. Four of these cultures were grown to an optical density at 600 nm (OD_{600 nm}) of 0.3, at which point induction of genes on pDB1282 and pBAD42-BtuCEDFB was initiated with the addition of arabinose to a final concentration of 0.2%. To facilitate iron–sulfur cluster incorporation, 25 µM FeCl₃ and 150 µM cysteine were added at the initiation of induction with arabinose. The cultures were then grown to an OD_{600 nm} of 0.6 followed by incubation in an ice-water bath for 1 h. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were incubated at 18 °C for an additional 18 h before the cells were collected by centrifugation at 6,000g. The resulting cell paste (around 50 g) was flash-frozen in liquid N₂ and stored in liquid N₂ before protein purification.

All purification steps and subsequent manipulations of TokK were performed in a Coy Laboratory Products vinyl anaerobic chamber. Cell paste was resuspended in 100 ml lysis buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 10% glycerol, 5 mM imidazole and 10 mM β-mercaptoethanol (BME)). The cell suspension was incubated with lysozyme (1 mg ml⁻¹), DNase (0.1 mg ml⁻¹) and phenylmethylsulfonyl fluoride (PMSF) (0.18 mg ml⁻¹) for 30 min at room temperature and then cooled to 4 °C (refs. 18,30). Cells were then lysed by sonic disruption (70% amplitude, 45 s on, 59 s off, around 15 min) and then centrifuged at 50,000g for 1 h to separate insoluble material. The resulting supernatant was loaded onto a pre-equilibrated column of Ni-NTA resin and purified by immobilized-metal affinity chromatography. The protein-loaded resin was washed with 50 ml lysis buffer before the addition of elution buffer (lysis buffer supplemented with 500 mM imidazole). Dark-coloured elution fractions were pooled

and concentrated to 1 ml in an Amicon 10 kDa MWCO ultrafiltration device (EMD Millipore). The protein fractions were then exchanged into TEV protease cleavage buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 15% glycerol and 10 mM BME) using a PD-10 pre-poured gel-filtration column from GE Biosciences. The exchanged protein sample was allowed to react with 10 units of TEV protease (around 2 mg ml⁻¹) in 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM TCEP and 50% glycerol (Millipore Sigma) for two days on ice to generate TokK containing only six additional amino acids (ENLYFQ) on its C terminus. On the second day, the [4Fe–4S] cluster and Cbl cofactors were reconstituted as previously described³⁰. The reaction mixture was then reapplied to the Ni-NTA resin to capture any remaining His-tagged protein, and purified TokK-ENLYFQ was collected in the flow-through fraction. TokK was concentrated and exchanged into gel-filtration buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 1 mM DTT and 15% glycerol) for size-exclusion chromatography. In this step, the protein was applied to a HiPrep 16/60 S200 column using an ÄKTA fast protein liquid chromatography (FPLC) system (GE Biosciences) housed in the anaerobic chamber. TokK elutes as an apparent monomer. Fractions were pooled on the basis of UV-vis absorption at 280 and 410 nm and concentrated to 38 mg ml⁻¹.

Synthesis of substrate

Synthesis of the TokK substrate, (2*R*,3*R*,5*R*)-3-((2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)thio)-7-oxo-1-azabicyclo[3.2.0]heptane-2-carboxylic acid (**1**), was carried out as previously described²⁷. Characterization matched that previously reported.

Determination of the X-ray crystal structure of TokK

General crystallographic methods

X-ray diffraction datasets were collected at the General Medical Sciences and Cancer Institutes Collaborative Access Team (GM/CA-CAT) and at the Advanced Photon Source, Argonne National Laboratory and Berkeley Center for Structural Biology (BCSB) beamlines at the Advanced Light Source at Lawrence Berkeley National Laboratory. All datasets were processed using the HKL2000 or HKL3000 package, and structures were determined by single anomalous dispersion (SAD) phasing using Autosol/HySS or by molecular replacement using the program PHASER^{31,32,33,34}. Model building and refinement were performed with Coot and phenix.refine^{31,35}. Figures were prepared using PyMOL^{36,37}. Substrate channel figures were prepared using Hollow³⁸. Ligplot was used to visualize the binding of substrate **1** (ref. ³⁹).

Crystallization and structure solution of 5'-dAH-Met-TokK

Purified TokK protein aliquots were diluted to 8 mg ml⁻¹ in 34 mM HEPES, pH 7.5. Then 5'-dAH and Met (Millipore Sigma) were added to the resulting solution to final concentrations of 2 mM each. The mixture was incubated for 30 min at room temperature. In hanging-drop vapour-diffusion trials with 100 mM magnesium chloride, 100 mM calcium chloride, 20% PEG 8000 and 10% 1,6-hexanediol as the precipitating reagent, brown plate-shaped crystals appeared within three days. Trials were initiated by adding 1 µl of protein with 1 µl precipitating solution, followed by equilibration against 500 µl of a 0.5 M LiCl well solution at room temperature. Crystals were prepared for data collection by mounting on rayon loops followed by a brief soak in cryoprotectant solution (50% (v/v) precipitating reagent and 50% (v/v) ethylene glycol) and flash-freezing in liquid nitrogen.

Diffraction datasets for single-wavelength anomalous diffraction phasing were collected at the iron K-edge X-ray absorption peak ($\lambda = 1.72194 \text{ \AA}$) with 360° of data measured using a 0.5° oscillation range to 2.52 Å resolution. In addition, a 1.79 Å-resolution native dataset was collected at $\lambda = 1.03313 \text{ \AA}$ (Extended Data Table 1). Heavy-atom sites were identified using HySS implemented within Phenix Autosol³¹. The initial overall figure-of-merit (FOM) was 0.265 and the Bayes CC was 18.2 (ref. 31). Phenix Autobuild was used to generate an initial model of 541 residues out of 687 in chain A and 569 residues out of 687 in chain B with $R_{\text{work}}/R_{\text{free}}$ of 0.25/0.31. Iterative manual model building and refinement were performed in Coot and Phenix³⁵. This model was used to obtain phase information for the 1.79 Å-resolution native dataset by molecular replacement using Phenix Phaser-MR³¹. Geometric restraints for 5'-dAH and Cbl were obtained from the Grade Web Server (Global Phasing). R_{free} flags were maintained so that the same 5% of the reflections were used as the test set. The final model consists of residues 9–412, 416–672, one [4Fe–4S] cluster, one OHCbl cofactor, one Met and one 5'-dAH in chain A; residues 7–413, 416–672, one [4Fe–4S] cluster, one OHCbl cofactor, one Met and one 5'-dAH in chain B. The final model also contains 2 chloride ions, 2 potassium ions, 21 molecules of ethylene glycol and 887 waters. The Ramachandran plot shows that 97.4% residues are in favoured regions with the remaining 2.6% in allowed regions³⁶. Data collection and refinement statistics are shown in Extended Data Table 1.

Crystallization and structure solution of substrate-bound 5'-dAH–Met–TokK

Purified TokK was diluted to 8 mg ml⁻¹ TokK in 34 mM HEPES, pH 7.5. Then 5'-dAH, Met and substrate were added to final concentrations of 2 mM each. The solution was incubated for 15 min at room temperature. In hanging-drop vapour diffusion crystallization trials with 0.2 M lithium sulfate, 0.1 M Tris-HCl, pH 8.5 and 18% PEG 8000 as the precipitant, brown plate-shaped crystals appeared within two weeks. Trials were initiated by mixing 1 µl of protein and 1 µl of precipitant followed

by equilibration at room temperature against a 500- μ l reservoir of the precipitating solution. Before looping, the concentration of substrate **1** was increased to 2.7 mM. Crystals were prepared for data collection by mounting on rayon loops followed by a brief soak in perfluoropolyether oil (Hampton Research) for cryoprotection and flash-freezing in liquid nitrogen.

The structure containing substrate was solved by molecular replacement (Phaser-MR in Phenix) using the coordinates of the 5'-dAH–Met–TokK as the search model. Iterative manual model building and refinement were performed in Coot and Phenix^{31,35}. Initial geometric restraints for (2*R*)-pantetheinylated carbapenem substrate were generated by eLBOW in Phenix³¹. Final geometric restraints for the carbapenem substrate were created by using the PRODRG 2 server⁴⁰. The final model consists of residues 8–566, 571–672, one [4Fe–4S] cluster, one OHcbl cofactor, one Met, one 5'-deoxyadenosine (5'-dAH) and one (2*R*)-pantetheinylated carbapenem substrate in chain A; residues 9–672, one [4Fe–4S] cluster, one OHcbl factor, one Met and one 5'-dAH in chain B. The final structure also contains 2 glycerol molecules, 4 potassium ions and 1,193 waters. The Ramachandran plot showed that 97.7% of residues are in favoured regions with the remaining 2.3% in allowed regions³⁶. Data collection and refinement statistics are shown in Extended Data Table [1](#).

Cbl ligand assignment in protein crystals

The protocol was adapted from previously published studies²⁸. Approximately 15 TokK + 5'-dAH + Met crystals were looped from the crystallization drop into 40 μ l of mother solution in a darkly coloured Eppendorf tube. In the dark, 50 mM H₂SO₄ was added to the tube. The tube was vortexed and then centrifuged for 20 min. Five microlitres of this solution was injected into a Thermo Fisher Scientific UHPLC/QExactive HF-X mass spectrometer equipped with a C18 column (2.1 \times 100 mm) equilibrated in 5% solvent A (0.1% formic acid) and 95% solvent B (0.1% formic acid in acetonitrile). The solvent B composition was increased to 98% from 1 to 7 min. Cbl forms were detected by ESI⁺, scanning from *m/z* 150 to 1,700 with a resolution of 120,000. A calibration curve (0.1 μ M–5 μ M) of Cbl standards was run concurrently to quantify the Cbl forms in the sample.

Generation of SSNs

A pool of annotated Cbl-dependent RS enzymes was generated by merging representative annotated B₁₂-dependent RS enzyme sequences from radicalsam.org (megacluster 2-1, subgroup 5; <https://radicalسام.org/explore.php?id=cluster-2-1&v=3.0>, accessed March 2021) with those from the USCF structure-function linkage database (SFLD) (<http://sfld.rbvi.ucsf.edu/archive/django/index.html>, accessed April

2021)^{41,42}. Additional sequences were included for the following functionally characterized enzymes (with Uniprot accession codes): *STsrM* (C0JRZ9), *KsTsrM* (E4N8S5), *GenK* (Q70KE5), *GenD1* (Q2MG55), *PhpK* (A0A0M3N271), *Fom3* (Q56184), *CysS* (A0A0H4NV78), *OxsB* (O24770), *PoyC* (J9ZXD6), *swb7* (D2KTX6), *ArgMT* (Q8THG6), *ThnK* (F8JND9), *TokK* (A0A6B9HEI0), *BchE* (Q7X2C7), *CouN6* (A0A1H2F7M3) and *CloN6* (Q9F8U1)^{7,12,14,15,18,43,44,45,46,47,48,49,50,51}. The pool of B_{12} -dependent RS enzyme sequences from radicalsam.org contains 9,724 representative entries that reflect 53,470 unique sequences. The class B and B_{12} -binding RS enzyme sequence groups from the SFLD contain 1,525 and 5,920 representatives, respectively. Each group reflects a total pool of 4,232 and 15,983 unique sequences, respectively⁴¹. The final representative sequence set used for SSN generation contained approximately 11,000 sequences after removal of duplicates. This dataset represents a larger sequence pool of more than 50,000 unique entries.

The Enzyme Function Initiative enzyme similarity tool (EFI-EST) (<https://efi.igb.illinois.edu>) was used to perform an all-by-all BLAST analysis of the representative sequence dataset described above to create an initial SSN^{42,52,53,54} with an alignment score threshold of 65. To eliminate protein fragments, sequence length was restricted to greater than 300 amino acids. The final SSN (Fig. 1c, Supplementary Fig. 1) is depicted as a representative node network in which each node reflects sequences with more than 40% identity. All networks were visualized with the Organic layout in Cytoscape⁵⁵. The SSN in Fig. 1c represents 36 sequence clusters extracted from the full SSN shown in Supplementary Fig. 1. The clusters in Fig. 1c were selected on the basis of the number of nodes or the presence of functionally annotated or structurally characterized sequences.

Preparation of EPR samples

TokK was diluted to a final concentration of approximately 250 μM for each EPR sample. The samples were photolysed on ice for 45 min to generate the cob(II)alamin state. All samples were flash-frozen in cryogenic liquid isopentane in an anaerobic chamber. The resulting samples were stored in liquid nitrogen before analysis. EPR measurements were taken on a Magnettech 5000 x-band ESR spectrometer equipped with an ER 4102ST resonator. Temperature was controlled by an ER 4112-HV Oxford Instruments variable-temperature helium-flow cryostat. All measurements were taken at 70 K, with 1 mT modulation amplitude and 1 mW power.

Construction of *TokK* variants

TokK variants were generated by overlap extension PCR using pET29b:*tokK*Tev as a template and the primers described in Supplementary Table 1. TokK_F_s was used as the forward primer for all constructs except the E19A/Y20V double variant, for which TokK_F_EYmut was used instead. After amplification, PCR products were digested with NdeI and XhoI and ligated into a similarly digested pET29b vector. Sequence-verified constructs were used to transform *E. coli* BL21 (DE3) along with helper plasmids pDB1282 and pBAD42-BtuCEDFB as described above for overexpression.

Methylation assays of TokK variants

Expression and purification of TokK variants was carried out as previously described for wild-type TokK¹⁸, except reconstitution³⁰ was done concurrently with overnight TEV protease cleavage, and an additional buffer exchange was done using an Econo-Pac 10DG column (Bio-Rad) to remove excess reconstitution reagents. Methylation assays were carried out in triplicate and contained 100 mM HEPES, pH 7.5, 200 mM KCl, 1 mM SAM, 1 mM methyl viologen, 2 mM NADPH, 0.5 mM MeCbl, 100 μM substrate and 100 μM enzyme. At each time point, an aliquot of the reaction mixture was diluted 5×, filtered through a 10-kDa Amicon ultrafiltration device and analysed for product formation using ultraperformance liquid chromatography-high resolution mass spectrometry (UPLC-HRMS) as previously described¹⁸. First-order rate constants were determined using the COPASI parameter estimation tool³⁶, and curves were simulated using Vcell³⁷.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the Protein Data Bank (PDB) under accession numbers [7KDX](#) (structure with 5'-dAH + Met) and [7KDY](#) (structure with 5'-dAH + Met + carbapenam substrate).

References

1. US Food and Drug Administration. Supplemental New Drug Approval: RECARBRIQ (imipenem, cilastatin, and relebactam) for injection, <https://www.fda.gov/RegulatoryInformation/Guidances/default.htm> (2020).

2. Fisher, J. F. & Mobashery, S. Three decades of the class A β -lactamase acyl-enzyme. *Curr. Protein Pept. Sci.* **10**, 401–407 (2009).
3. Sinner, E. K., Marous, D. R. & Townsend, C. A. Evolution of methods for the study of cobalamin-dependent radical SAM enzymes. *ACS Bio. Med. Chem. Au* <https://doi.org/10.1021/acsbiomedchemau.1c00032> (2021).
4. Li, R., Lloyd, E. P., Moshos, K. A. & Townsend, C. A. Identification and characterization of the carbapenem MM 4550 and its gene cluster in *Streptomyces argenteolus* ATCC 11009. *ChemBioChem* **15**, 320–331 (2014).
5. Nunez, L. E., Mendez, C., Brana, A. F., Blanco, G. & Salas, J. A. The biosynthetic gene cluster for the β -lactam carbapenem thienamycin in *Streptomyces cattleya*. *Chem. Biol.* **10**, 301–311 (2003).
6. Coulthurst, S. J., Barnard, A. M. & Salmond, G. P. Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nat. Rev. Microbiol.* **3**, 295–306 (2005).
7. Wang, Y., Schnell, B., Baumann, S., Muller, R. & Begley, T. P. Biosynthesis of branched alkoxy groups: iterative methyl group alkylation by a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **139**, 1742–1745 (2017).
8. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
9. Holliday, G. L. et al. Atlas of the radical SAM superfamily: divergent evolution of function using a “plug and play” domain. *Methods Enzymol.* **606**, 1–71 (2018).
10. Wang, S. C. Cobalamin-dependent radical S-adenosyl-l-methionine enzymes in natural product biosynthesis. *Nat. Prod. Rep.* **35**, 707–720 (2018).
11. Miller, D. V. et al. in *Comprehensive Natural Products III: Chemistry and Biology* vol. 5 (eds Liu, H.-W. & Begley, T.) 24–69 (Elsevier, 2020).
12. Knox, H. L. et al. Structural basis for non-radical catalysis by TsrM, a radical SAM methylase. *Nat. Chem. Biol.* **17**, 485–491 (2021).
13. Pierre, S. et al. Thiomodulon tryptophan methyltransferase expands the chemistry of radical SAM enzymes. *Nat. Chem. Biol.* **8**, 957–959 (2012).
14. Blaszczyk, A. J. et al. Spectroscopic and electrochemical characterization of the iron-sulfur and cobalamin cofactors of TsrM, an unusual radical S-adenosylmethionine methylase. *J. Am. Chem. Soc.* **138**, 3416–3426 (2016).

15. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L. & Liu, H. W. A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature* **544**, 322–326 (2017).
16. Blaszczyk, A. J., Wang, B., Silakov, A., Ho, J. V. & Booker, S. J. Efficient methylation of C2 in l-tryptophan by the cobalamin-dependent radical *S*-adenosylmethionine methylase TsrM requires an unmodified N1 amine. *J. Biol. Chem.* **292**, 15456–15467 (2017).
17. Chang, W.-C. et al. Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science* **343**, 1140–1144 (2014).
18. Sinner, E. K., Lichstrahl, M. S., Li, R., Marous, D. R. & Townsend, C. A. Methylations in complex carbapenem biosynthesis are catalyzed by a single cobalamin-dependent radical *S*-adenosylmethionine enzyme. *Chem. Commun.* **55**, 14934–14937 (2019).
19. Albers-Schoenberg, G. A. et al. Structure and absolute configuration of thienamycin. *J. Am. Chem. Soc.* **100**, 6491–6499 (1978).
20. Mitchell, A. J. et al. Visualizing the reaction cycle in an iron(II)- and 2-(oxo)-glutarate-dependent hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
21. Poulos, T. L., Finzel, B. C. & Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **195**, 687–700 (1987).
22. Schwalm, E. L., Grove, T. L., Booker, S. J. & Boal, A. K. Crystallographic capture of a radical *S*-adenosylmethionine enzyme in the act of modifying tRNA. *Science* **352**, 309–312 (2016).
23. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a protein binds B-12—a 3.0-Å X-ray structure of B-12-binding domains of methionine synthase. *Science* **266**, 1669–1674 (1994).
24. Kräutler, B. Thermodynamic *trans*-effects of the nucleotide base in the B12 coenzymes. *Helv. Chim. Acta* **70**, 1268–1278 (1987).
25. Krautler, B. Chemistry of methylcorrinoids related to their roles in bacterial C1 metabolism. *FEMS Microbiol. Rev.* **7**, 349–354 (1990).
26. Milton, P. A. & Brown, T. L. The kinetics of benzimidazole dissociation in methylcobalamin. *J. Am. Chem. Soc.* **99**, 1390–1396 (1977).

27. Marous, D. R. et al. Consecutive radical *S*-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis. *Proc. Natl Acad. Sci USA* **112**, 10354–10358 (2015).
28. Lanz, N. D. et al. Enhanced solubilization of class B radical *S*-adenosylmethionine methylases by improved cobalamin uptake in *Escherichia coli*. *Biochemistry* **57**, 1475–1490 (2018).
29. Lanz, N. D. et al. RIMN and AtsB as models for the overproduction and characterization radical SAM Proteins. *Methods Enzymol.* **516**, 125–152 (2012).
30. Blaszczyk, A. J., Wang, R. X. & Booker, S. J. TsrM as a model for purifying and characterizing cobalamin-dependent radical *S*-adenosylmethionine methylases. *Methods Enzymol.* **595**, 303–329 (2017).
31. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
32. Bunkoczi, G. et al. Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D* **69**, 2276–2286 (2013).
33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
36. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
37. The PyMOL Molecular Graphics Systems v.2.0 (Schrödinger).
38. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
39. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).

40. Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
41. Akiva, E. et al. The Structure–Function Linkage Database. *Nucleic Acids Res.* **42**, D521–D530 (2014).
42. Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* **56**, 4293–4308 (2017).
43. Kim, H. J. et al. GenK-catalyzed C-6' methylation in the biosynthesis of gentamicin: isolation and characterization of a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **135**, 8093–8096 (2013).
44. Huang, C. et al. Delineating the biosynthesis of gentamicin x2, the common precursor of the gentamicin C antibiotic complex. *Chem. Biol.* **22**, 251–261 (2015).
45. Werner, W. J. et al. In vitro phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry* **50**, 8986–8988 (2011).
46. Wang, B. et al. Stereochemical and mechanistic investigation of the reaction catalyzed by Fom3 from *Streptomyces fradiae*, a cobalamin-dependent radical S-adenosylmethionine methylase. *Biochemistry* **57**, 4972–4984 (2018).
47. Parent, A. et al. The B12-radical SAM enzyme PoyC catalyzes valine C β -methylation during polytheonamide biosynthesis. *J. Am. Chem. Soc.* **138**, 15515–15518 (2016).
48. Watanabe, K. et al. *Escherichia coli* allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product. *J. Am. Chem. Soc.* **131**, 9347–9353 (2009).
49. Suzuki, J. Y., Bolivar, D. W. & Bauer, C. E. Genetic analysis of chlorophyll biosynthesis. *Annu. Rev. Genet.* **31**, 61–89 (1997).
50. Westrich, L., Heide, L. & Li, S. M. CloN6, a novel methyltransferase catalysing the methylation of the pyrrole-2-carboxyl moiety of clorobiocin. *ChemBioChem* **4**, 768–773 (2003).
51. Radle, M. I., Miller, D. V., Laremore, T. N. & Booker, S. J. Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical S-

- adenosylmethionine methylase that unexpectedly requires cobalamin. *J. Biol. Chem.* **294**, 11712–11725 (2019).
52. Gerlt, J. A. et al. Enzyme Function Initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
53. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182 (2019).
54. Zallot, R., Oberg, N. O. & Gerlt, J. A. ‘Democratized’ genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **47**, 77–85 (2018).
55. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
56. Hoops, S. et al. COPASI-a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067–3074 (2006).
57. Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H. & Loew, L. M. A general computational framework for modeling cellular structure and function. *Biophys. J.* **73**, 1135–1146 (1997).

Acknowledgements

This work was supported by the National Institutes of Health (NIH) (GM122595 to S.J.B., GM119707 to A.K.B., AI121072 to C.A.T. and GM080189 to E.K.S.) and the Eberly Family Distinguished Chair in Science (S.J.B.). S.J.B. is an investigator of the Howard Hughes Medical Institute. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of GM/CA@APS has been funded in whole or in part with federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). This research also used the resources of the Berkeley Center for Structural Biology supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under contract no. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the NIH, National Institute of

General Medical Sciences, grant P30 GM124169. E.K.S. and C.A.T. thank M. S. Lichstrahl for providing a synthetic intermediate, and I. P. Mortimer and J. Tang for their help with ESI-MS and NMR experiments, respectively. We thank D. Iwig for his help in the mass spectrometry analysis of the TokK crystals.

Author information

Author notes

1. These authors contributed equally: Hayley L. Knox, Erica K. Sinner

Affiliations

1. Department of Chemistry, Pennsylvania State University, University Park, PA, USA

Hayley L. Knox, Amie K. Boal & Squire J. Booker

2. Department of Chemistry, Johns Hopkins University, Baltimore, MD, USA

Erica K. Sinner & Craig A. Townsend

3. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA

Amie K. Boal & Squire J. Booker

4. The Howard Hughes Medical Institute, Pennsylvania State University, University Park, PA, USA

Squire J. Booker

Contributions

H.L.K., E.K.S., C.A.T. and S.J.B. developed the research plan and experimental strategy. H.L.K. isolated and crystallized proteins and collected crystallographic data. H.L.K. and E.K.S. performed biochemical experiments. H.L.K., E.K.S., C.A.T., A.K.B. and S.J.B. analysed and interpreted crystallographic data. H.L.K., E.K.S., C.A.T., A.K.B. and S.J.B. wrote the manuscript.

Corresponding authors

Correspondence to [Craig A. Townsend](#), [Amie K. Boal](#) or [Squire J. Booker](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

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 Comparison to CysS, another class B sequential methylase.](#)

a, CysS performs sequential radical methylations like TokK and ThnK. **b**, Partial alignment of TokK (Uniprot ID: A0A6B9HEI0), ThnK (Uniprot ID: F8JND9), CysS (Uniprot ID: A0A0H4NV78), TsrM (Uniprot ID: C0JRZ9), Fom3 (Uniprot ID: Q56184), PhpK (Uniprot ID: A0A0M3N271), and GenK (Uniprot ID: Q70KE5). Cysteines that coordinate the iron-sulfur cluster are highlighted in blue. Trp76 (highlighted in red) and Trp215 (highlighted in yellow) in TokK is conserved in ThnK and CysS, but not in other known Cbl-binding RS methylases. Areas of conservation for TokK, ThnK, and CysS around Trp76 are highlighted in grey. The bottom axial amino acid residue for TsrM (Arg69) is highlighted in red. Completely conserved residues are bolded.

[Extended Data Fig. 2 Reactions performed by OxsB and TsrM, notable Cbl-binding radical SAM enzymes.](#)

a, Proposed pathway for the biosynthesis of oxetanocin A by OxsB and OxsA. Currently the aldehyde reduction step is unknown. **b**, Proposed non-radical reaction performed by TsrM. The carboxylate in SAM is implicated in playing a dual role in catalysis, both as the source of the methyl donor and as the base to prime the substrate for nucleophilic attack.

Extended Data Fig. 3 Structural comparisons of Cbl-dependent RS enzymes.

a–c, The overall structures of TokK (**a**), *KsTsrM* (PDB ID: 6WTE) (**b**) and OxsB (PDB ID: 5UL3) (**c**) are shown as ribbon diagrams and coloured by domain (Cbl-binding domain, teal; RS domain, light blue; and C-terminal domain, pink). OxsB has a fourth domain, an N-terminal domain of an unknown function, shown in gold. All three enzymes share very similar Cbl-binding domains with a characteristic Rossmann fold. However, as shown in Extended Data Fig. 4, the RS domain and C-terminal domain differ in each system.

Extended Data Fig. 4 Comparison of the Cbl-binding, RS, and C-terminal domains of TokK, KsTsrM, and OxsB.

The domains are coloured as in Extended Data Fig. 3. **a**, The Rossmann fold, in teal, is highly similar among TokK (PDB ID: 7KDY), *KsTsrM* (PDB ID: 6WTF) and OxsB (PDB ID: 5UL4). **b**, The core of each of the RS domains is a $(\beta/\alpha)_6$ motif; however, there are distinct differences. The RS domain of OxsB is more compact than those of TokK or *KsTsrM*, and all three have unique extra secondary structure features. **c**, The C-terminal domains for TokK, *KsTsrM*, and OxsB are vastly different in architecture. **d**, Comparison of the binding of Met and 5'-dAH, aza-SAM, and SAM for TokK, *KsTsrM*, and OxsB structures, respectively. Only the relevant binding of SAM to the cluster is shown for OxsB. OxsB has two binding positions of SAM, one to the cluster and one in what is proposed to be an intermediate state towards methylating the Cbl.

Extended Data Fig. 5 Domain architecture of TokK.

a, The three domains of TokK are portrayed in a ribbon diagram. The N-terminal Cbl-binding domain is shown in teal, the RS domain is shown in light blue, and the C-terminal domain is shown in pink. **b**, A topology diagram of TokK with domains coloured as in **a**. The Cbl is portrayed in sticks, and the lower axial Trp side chain is shown as a red dot. The iron-sulfur cluster is shown in orange and yellow spheres, and the ligating Cys residues are shown as small yellow spheres. **c–e**, Zoomed in views of the C-terminal domain (**c**), RS domain (**d**) and Cbl-binding domain (**e**) are shown as ribbon diagrams.

Extended Data Fig. 6 Diagram of carbapenam interactions in the TokK complex with substrate.

Carbapenam substrate **1** shown in blue sticks and coloured by atom type. Polar and hydrophobic interactions were mapped with the LigPlot program and indicated with dashed lines or a starburst symbol.

Extended Data Fig. 7 Comparison of TokK structure with those of enzymes using alternative radical-generating mechanisms.

a–e, Residues that interact with the polar substituents in the β -lactam ring are shown in stick format for TokK (**a**) and the epimerase CarC (PDB ID: 4OJ8) (**b**). Although the two enzymes are structurally distinct, they use a similar number and type of functional groups to anchor the β -lactam by using direct and water-mediated contacts to the C7 carbonyl and C3 carboxylate substituents. In addition, comparison of the active site with those of hydroxylases reveals differences in the orientation of hydrogen atom transfer (HAT) intermediates and -OH or -CH₃ functionalization moieties. Fe(II)- and 2-oxo-glutarate-dependent (Fe-2OG) oxygenases use a ferryl [Fe(IV)-oxo] intermediate to abstract an H-atom from an unactivated substrate carbon. The resulting Fe(III)-OH complex then couples with the substrate radical to yield a hydroxylated product. A vanadyl [V(IV)-oxo] mimic of the ferryl intermediate in L-Arg C3 hydroxylase, VioC, reveals that HAT and -OH transfer must occur from the same side of the substrate target carbon (**c**). The distance between the reactive oxo group and the substrate target carbon (indicated by the arrows) is 3.1 Å in VioC. A structure of the haem-dependent hydroxylase, P450cam, shows a similar phenomenon. A CO-bound mimic of the Fe(IV)-porphyrin radical intermediate, compound I, demands the same arrangement of HAT and -OH transfer components relative to the hydroxylation target on the camphor substrate (**d**). The distance between a mimic of the reactive group and the substrate target carbon (indicated by the arrows) is 3.1 Å in P450cam. By contrast, RS methylases use separate HAT reagents (5'-dA \bullet) and methyl group donors (Me-Cbl), allowing for more diverse stereochemical outcomes in C-H functionalization reactions. (**e**) The structure of TokK in complex with carbapenam substrate, **1**, shows a ~120° angle between the HAT acceptor (5'C of 5'-dAH), the substrate target carbon (C6), and the Cbl top ligand (-OH of OHcbl, a surrogate for Me-Cbl).

Extended Data Fig. 8 A comparison of the substrate-binding sites in two structurally characterized RS methylases.

a, b, The substrate complexes of TokK (**a**) and RlmN (PDB ID: 5HR7) (**b**) are shown. RlmN is an RS methylase that uses a radical-based mechanism to methylate an *sp*²-hybridized carbon (C2) of an adenine base in transfer or ribosomal RNA. By contrast

to TokK and other Cbl-dependent RS enzymes, RlmN uses a 5'-dA[•] to activate a post-translationally modified methyl-Cys residue on a C-terminal loop in the active site to modify its aromatic substrate via radical addition. Comparison of a structure of RlmN in a cross-linked methylCys-tRNA intermediate state to the TokK substrate complex shows that, despite the differences in reaction mechanism, these systems use a similar orientation of radical initiator (5'-dA[•]), substrate target carbon (C6, C2), and methyl donor (OH, Me). In the TokK substrate complex, the -OH ligand of the Cbl cofactor serves as a surrogate for the position of the methyl donor.

Extended Data Fig. 9 Solvent accessibility of the bottom axial face of the Cbl in TokK.

a, Nearest residues (~6 Å) to the bottom face of the Cbl. **b**, View of TokK using space-filling model to show a channel to the active site. As can be seen, only a very small portion of the Cbl (coloured yellow) is solvent accessible, with most of the Cbl being buried within the Rossmann fold. Trp76 is coloured red.

Extended Data Table 1 X-ray crystallographic data collection and refinement statistics

Supplementary information

Supplementary Information

This file contains Supplementary Figures 1–7 and Supplementary Table 1.

Reporting Summary

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Knox, H.L., Sinner, E.K., Townsend, C.A. *et al.* Structure of a B₁₂-dependent radical SAM enzyme in carbapenem biosynthesis. *Nature* **602**, 343–348 (2022).
<https://doi.org/10.1038/s41586-021-04392-4>

- Received: 10 July 2021
- Accepted: 22 December 2021
- Published: 02 February 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04392-4>

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/s41586-021-04392-4>

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

Amendments & Corrections

- [**Author Correction: Coupling of Indo-Pacific climate variability over the last millennium**](#) [12 January 2022]
Author Correction •
- [**Author Correction: A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types**](#) [12 January 2022]
Author Correction •
- [**Author Correction: Unextractable fossil fuels in a 1.5 °C world**](#) [25 January 2022]
Author Correction •
- [**Publisher Correction: Optomechanical dissipative solitons**](#) [24 January 2022]
Publisher Correction •

- Author Correction
- [Published: 12 January 2022](#)

Author Correction: Coupling of Indo-Pacific climate variability over the last millennium

- [Nerilie J. Abram](#) ORCID: [orcid.org/0000-0003-1246-2344^{1,2}](https://orcid.org/0000-0003-1246-2344),
- [Nicky M. Wright](#) ORCID: [orcid.org/0000-0002-5600-3193^{1,2}](https://orcid.org/0000-0002-5600-3193),
- [Bethany Ellis](#) ORCID: [orcid.org/0000-0002-4662-1115^{1,3}](https://orcid.org/0000-0002-4662-1115),
- [Bronwyn C. Dixon^{1,3,4}](#),
- [Jennifer B. Wurtzel^{1 nAff15}](#),
- [Matthew H. England](#) ORCID: [orcid.org/0000-0001-9696-2930^{5,6}](https://orcid.org/0000-0001-9696-2930),
- [Caroline C. Ummenhofer](#) ORCID: [orcid.org/0000-0002-9163-3967^{6,7}](https://orcid.org/0000-0002-9163-3967),
- [Belle Philibosian⁸](#),
- [Sri Yudawati Cahyarini⁹](#),
- [Tsai-Luen Yu^{10,11}](#),
- [Chuan-Chou Shen^{10,11,12}](#),
- [Hai Cheng^{13,14}](#),
- [R. Lawrence Edwards¹⁴](#) &
- [David Heslop¹](#)

[Nature](#) volume **602**, page E20 (2022)

- 662 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Climate change](#)
- [Palaeoclimate](#)

The [Original Article](#) was published on 09 March 2020

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-020-2084-4> Published online 9 March 2020

There are two amendments to the version of this article initially published. The Fig. 2 caption has been updated to specify the data sources described in the Methods. Further, an update has been made to Supplementary Figure 10b–e to correct an error that resulted from a mismatch of units between centimetres and millimetres in the original code. We also note that this simulation tested sampling of a coral with linear extension rate of 20 mm y^{-1} across an extreme range of edge thicknesses from 1–7 mm, but that the realistic range of edge thicknesses for our study is 2.5–4 mm.

The changes have been made to the online version of the article.

Author information

Author notes

1. Jennifer B. Wurtzel

Present address: New South Wales Department of Primary Industries, Orange, New South Wales, Australia

Affiliations

1. Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory, Australia

Nerilie J. Abram, Nicky M. Wright, Bethany Ellis, Bronwyn C. Dixon, Jennifer B. Wurtzel & David Heslop

2. ARC Centre of Excellence for Climate Extremes, The Australian National University, Canberra, Australian Capital Territory, Australia

Nerilie J. Abram & Nicky M. Wright

3. ARC Centre of Excellence for Climate System Science, The Australian National University, Canberra, Australian Capital Territory, Australia

Bethany Ellis & Bronwyn C. Dixon

4. School of Geography, University of Melbourne, Melbourne, Victoria, Australia

Bronwyn C. Dixon

5. Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia

Matthew H. England

6. ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia

Matthew H. England & Caroline C. Ummenhofer

7. Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Caroline C. Ummenhofer

8. Earthquake Science Center, United States Geological Survey, Menlo Park, CA, USA

Belle Philibosian

9. Research Centre of Geotechnology, Indonesian Institute of Sciences (LIPI), Bandung, Indonesia

Sri Yudawati Cahyarini

10. High-precision Mass Spectrometry and Environment Change Laboratory (HISPEC), Department of Geosciences, National Taiwan University, Taipei, Taiwan

Tsai-Luen Yu & Chuan-Chou Shen

11. Research Center for Future Earth, National Taiwan University, Taipei, Taiwan

Tsai-Luen Yu & Chuan-Chou Shen

12. Global Change Research Center, National Taiwan University, Taipei, Taiwan

Chuan-Chou Shen

13. Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an, China

Hai Cheng

14. Department of Geology and Geophysics, University of Minnesota, Minneapolis, MN, USA

Hai Cheng & R. Lawrence Edwards

Corresponding author

Correspondence to [Nerilie J. Abram](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Abram, N.J., Wright, N.M., Ellis, B. *et al.* Author Correction: Coupling of Indo-Pacific climate variability over the last millennium. *Nature* **602**, E20 (2022). <https://doi.org/10.1038/s41586-021-04318-0>

- Published: 12 January 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04318-0>

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/s41586-021-04318-0>

- Author Correction
- Open Access
- [Published: 12 January 2022](#)

Author Correction: A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types

- [Velina Kozareva¹](#) na1,
- [Caroline Martin¹](#) na1,
- [Tomas Osorno²](#),
- [Stephanie Rudolph²](#),
- [Chong Guo](#) [ORCID: orcid.org/0000-0002-1230-5333²](#),
- [Charles Vanderburg¹](#),
- [Naeem Nadaf¹](#),
- [Aviv Regev](#) [ORCID: orcid.org/0000-0003-3293-3158¹](#),
- [Wade G. Regehr²](#) &
- [Evan Macosko](#) [ORCID: orcid.org/0000-0002-2794-5165^{1,3}](#)

Nature volume **602**, page E21 (2022)

- 851 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Cellular neuroscience](#)
- [Genomics](#)

The [Original Article](#) was published on 06 October 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03220-z> Published online 6 October 2021

In the version of this article initially published, there was a typo in the Data availability section. The Gene Expression Omnibus accession has been updated to [GSE165371](#).

The changes have been made in the online version of the paper.

Author information

Author notes

1. These authors contributed equally: Velina Kozareva, Caroline Martin

Affiliations

1. Broad Institute of Harvard and MIT, Stanley Center for Psychiatric Research, Cambridge, MA, USA

Velina Kozareva, Caroline Martin, Charles Vanderburg, Naeem Nadaf, Aviv Regev & Evan Macosko

2. Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Tomas Osorno, Stephanie Rudolph, Chong Guo & Wade G. Regehr

3. Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

Evan Macosko

Corresponding author

Correspondence to [Evan Macosko](#).

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

Kozareva, V., Martin, C., Osorno, T. *et al.* Author Correction: A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. *Nature* **602**, E21 (2022). <https://doi.org/10.1038/s41586-021-04373-7>

- Published: 12 January 2022

- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04373-7>

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/s41586-021-04373-7>

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

- Author Correction
- [Published: 25 January 2022](#)

Author Correction: Unextractable fossil fuels in a 1.5 °C world

- [Dan Welsby](#) ORCID: [orcid.org/0000-0002-8800-0229¹](https://orcid.org/0000-0002-8800-0229),
- [James Price²](#),
- [Steve Pye](#) ORCID: [orcid.org/0000-0003-1793-2552²](https://orcid.org/0000-0003-1793-2552) &
- [Paul Ekins¹](#)

[Nature](#) volume 602, pages E22–E23 (2022)

- 647 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Energy modelling](#)
- [Fossil fuels](#)

The [Original Article](#) was published on 08 September 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03821-8> Published online 08 September 2021.

In this Article, a small, unintentional off-model calculation oversight meant that the volumes of unextractable oil and fossil methane gas for some categories of oil and gas in some regions were overestimated. Having corrected for this oversight, at a global level, unextractable oil and coal reserves in 2050 have not changed, whereas fossil methane gas is 3% lower than in our published estimates. Here we describe the differences between the corrected and original published estimates. It should be noted that most of our analysis (production pathways and unextractable resource estimates) remain unchanged from the published Article. The correction affects light tight oil, shale gas, tight gas and coalbed methane in regions where these categories form a part of the proved reserve base, namely, Australia, Canada, China and the United States. The error arose because there is no explicit distinction between reserves and resources for these categories in the supply cost curves used in our model, TIAM-UCL.

Tables [1](#) and [2](#) (oil and fossil methane gas, respectively) of this Amendment show the corrected regional and global estimates of unextractable oil and gas. Alterations to the original Article are as follows: (1) the percentage of unextractable global fossil methane gas reserves in 2050 has been changed from 59% to 56%; (2) the percentage of unextractable oil has been corrected from 43% to 42% in 2100, and fossil methane gas has been corrected from 50% to 47%.

Table 1 Regional and global difference in unextractable oil reserves after accounting error has been fixed

Table 2 Regional and global difference in unextractable gas fossil methane reserves after accounting error has been fixed

Table [1](#) and Fig. 1 of the original Article have been corrected to reflect the revised estimates of unextractable oil and gas reserves for the regions impacted, as shown in Tables [1](#) and [2](#), respectively, of this Amendment. For reference, the corrected version of Table [1](#) is shown below as Table [3](#).

Table 3 Corrected Table [1](#) from the original Article

In addition, the Supplementary Information of the original Article has been altered to reflect the corrected estimates of unextractable oil and gas

reserves. The following changes were made to the Supplementary Information:

- A new section (section 8) explains in detail the source off the off-model calculation error and a step-by-step example for the United States.
- Supplementary Figs. 3 and 4 and Supplementary Tables 2 and 3 have been altered to reflect the corrected unextractable volumes (740 Gb and 87 tcm) of oil and gas and unextractable proportions of gas (56%). In addition, the text has been changed to reflect the reduction in unextractable global fossil methane gas reserves (from 59% to 56%) at the start of the section ‘Comparison to McGlade and Ekins (2015) paper’.
- A new file has been added that contains Source Data for Supplementary Figs. 1–16.

All authors agreed with this Author Correction, and we thank the peer reviewers for their feedback, particularly in communicating the correction in a clear and concise manner. The original Article has been corrected online

Author information

Affiliations

1. Institute for Sustainable Resources, University College London, London, UK

Dan Welsby & Paul Ekins

2. UCL Energy Institute, University College London, London, UK

James Price & Steve Pye

Corresponding author

Correspondence to [Dan Welsby](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Welsby, D., Price, J., Pye, S. *et al.* Author Correction: Unextractable fossil fuels in a 1.5 °C world. *Nature* **602**, E22–E23 (2022).

<https://doi.org/10.1038/s41586-021-04334-0>

- Published: 25 January 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04334-0>

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/s41586-021-04334-0>

- Publisher Correction
- [Published: 24 January 2022](#)

Publisher Correction: Optomechanical dissipative solitons

- [Jing Zhang^{1,2}](#),
- [Bo Peng¹](#),
- [Seunghwi Kim³](#),
- [Faraz Monifi¹](#),
- [Xuefeng Jiang¹](#),
- [Yihang Li¹](#),
- [Peng Yu⁴](#),
- [Lianqing Liu⁴](#),
- [Yu-xi Liu⁵](#),
- [Andrea Alù ORCID: orcid.org/0000-0002-4297-5274^{3,6}](#) &
- [Lan Yang ORCID: orcid.org/0000-0002-9052-0450¹](#)

[Nature](#) volume **602**, page E24 (2022)

- 591 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Nonlinear optics](#)
- [Silicon photonics](#)
- [Solitons](#)

The [Original Article](#) was published on 01 December 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04012-1> Published online 1 December 2021

In the version of this article initially published, there were errors in Fig. 2h, i. The y -axis labels for both panels, now reading “ $\omega_\mu - \Omega_m - D_1\mu$ (kHz),” were missing minus signs after “ Ω_m .”

The changes have been made to the online version of the article.

Author information

Affiliations

1. Department of Electrical and Systems Engineering, Washington University, St. Louis, MO, USA

Jing Zhang, Bo Peng, Faraz Monifi, Xuefeng Jiang, Yihang Li & Lan Yang

2. Department of Automation, Tsinghua University, Beijing, P. R. China

Jing Zhang

3. Photonics Initiative, Advanced Science Research Center, City University of New York, New York, NY, USA

Seunghwi Kim & Andrea Alù

4. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, P. R. China

Peng Yu & Lianqing Liu

5. Institute of Microelectronics, Tsinghua University, Beijing, P. R. China

Yu-xi Liu

6. Physics Program, Graduate Center, City University of New York, New York, NY, USA

Andrea Alù

Corresponding author

Correspondence to [Lan Yang](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Zhang, J., Peng, B., Kim, S. *et al.* Publisher Correction: Optomechanical dissipative solitons. *Nature* **602**, E24 (2022).
<https://doi.org/10.1038/s41586-021-04317-1>

- Published: 24 January 2022
- Issue Date: 10 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04317-1>

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/s41586-021-04317-1>

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