

# nature



## SHOOTING THE BREEZE

Seed-inspired microfliers  
hitch a lift on the wind

**Population boom**  
How many people will  
there be on Earth by  
the end of the century?

**Hiding in waiting**  
Return of Ebola to  
Guinea linked to latent  
virus in people

**Burning issue**  
The effect of fires  
and deforestation on  
Amazon biodiversity

# Nature.2021.09.25

[Sat, 25 Sep 2021]

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

# This Week

- **[Greece used AI to curb COVID: what other nations can learn](#)** [ 22 September 2021]  
Editorial • Governments are hungry to deploy big data in health emergencies. Scientists must help to lay the legal, ethical and logistical groundwork.
- **[Biden must keep funding pledge to historically Black colleges and universities](#)** [ 21 September 2021]  
Editorial • Congress has eviscerated the US president's pledge to upgrade research infrastructure at universities that serve underrepresented communities.
- **[Hurricane Ida shows the one-two punch of poverty and climate change](#)** [ 21 September 2021]  
World View • US President Joe Biden's environmental-justice adviser says: tackle inequality and global warming together.
- **[A CRISPR fix for muscles hatches from a viral shell](#)** [ 15 September 2021]  
Research Highlight • Scientists create millions of mutant viruses to find those that excel at ferrying genome-editing tools into muscle tissue.
- **[China faces one–two punches of extreme weather as Earth warms](#)** [ 15 September 2021]  
Research Highlight • The chances of heavy rain and high heat within one week are higher now than in the past two millennia, records suggest.
- **[An unruly painkiller is tamed with inspiration from nature](#)** [ 15 September 2021]  
Research Highlight • Scientists borrow features of the body's receptor for tetrodotoxin to create a useful synthetic structure.
- **[Home working brings longer hours, fewer phone calls](#)** [ 14 September 2021]  
Research Highlight • Data on more than 60,000 workers at Microsoft show that remote working led to more 'siloed' work groups and the sending of more e-mails.
- **[Humans walk efficiently even with their heads in the clouds](#)** [ 17 September 2021]  
Research Highlight • Exoskeleton-clad volunteers show that adapting to an energy-saving pace requires almost no attention.

- **A sweet tooth gave ancient primates a mouthful of woe** [ 14  
September 2021]  
Research Highlight • Fossils of a monkey-like animal that lived tens of millions of years ago furnish the earliest evidence of a mammal with cavities.
- **Puffins and friends suffer in washing-machine waves** [ 13  
September 2021]  
Research Highlight • Cyclones could make it difficult for seabirds such as little auks and puffins to hunt, which can lead to their starvation.

---

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

- EDITORIAL
- 22 September 2021

# Greece used AI to curb COVID: what other nations can learn

Governments are hungry to deploy big data in health emergencies. Scientists must help to lay the legal, ethical and logistical groundwork.



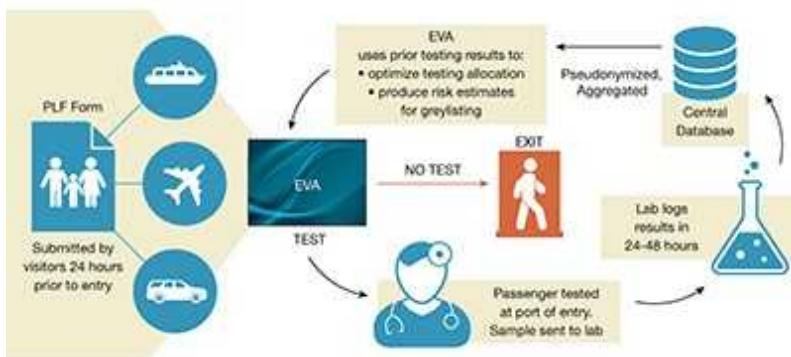
Greece's decision to deploy machine learning in pandemic surveillance will be much-studied around the world. Credit: Konstantinos Tsakalidis/Bloomberg/Getty

A few months into the COVID-19 pandemic, operations researcher Kimon Drakopoulos e-mailed both the Greek prime minister and the head of the country's COVID-19 scientific task force to ask if they needed any extra advice.

Drakopoulos works in data science at the University of Southern California in Los Angeles, and is originally from Greece. To his surprise, he received a reply from Prime Minister Kyriakos Mitsotakis within hours. The European Union was asking member states, many of which had implemented widespread lockdowns in March, to allow non-essential travel to recommence from July 2020, and the Greek government needed help in deciding when and how to reopen borders.

Greece, like many other countries, lacked the capacity to test all travellers, particularly those not displaying symptoms. One option was to test a sample of visitors, but Greece opted to trial an approach rooted in artificial intelligence (AI).

Between August and November 2020 — with input from Drakopoulos and his colleagues — the authorities launched a system that uses a machine-learning algorithm to determine which travellers entering the country should be tested for COVID-19. The authors found machine learning to be more effective at identifying asymptomatic people than was random testing or testing based on a traveller's country of origin. According to the researchers' analysis, during the peak tourist season, the system detected two to four times more infected travellers than did random testing.



[Read the paper: Efficient and targeted COVID-19 border testing via reinforcement learning](#)

The machine-learning system, which is among the first of its kind, is called Eva and is described in *Nature* this week ([H. Bastani et al. \*Nature\* <https://doi.org/10.1038/s41586-021-04014-z>; 2021](#)). It's an example of how data analysis can contribute to effective COVID-19 policies. But it also presents challenges, from ensuring that individuals' privacy is protected to the need to independently verify its accuracy. Moreover, Eva is a reminder of why proposals for a pandemic treaty (see [Nature 594, 8; 2021](#)) must consider rules and protocols on the proper use of AI and big data. These need to be drawn up in advance so that such analyses can be used quickly and safely in an emergency.

In many countries, travellers are chosen for COVID-19 testing at random or according to risk categories. For example, a person coming from a region with a high rate of infections might be prioritized for testing over someone travelling from a region with a lower rate.

By contrast, Eva collected not only travel history, but also demographic data such as age and sex from the passenger information forms required for entry to Greece. It then matched those characteristics with data from previously tested passengers and used the results to estimate an individual's risk of infection. COVID-19 tests were targeted to travellers calculated to be at highest risk. The algorithm also issued tests to allow it to fill data gaps, ensuring that it remained up to date as the situation unfolded.

During the pandemic, there has been no shortage of ideas on how to deploy big data and AI to improve public health or assess the pandemic's economic impact. However, relatively few of these ideas have made it into practice. This is partly because companies and governments that hold relevant data — such as mobile-phone records or details of financial transactions — need agreed systems to be in place before they can share the data with researchers. It's also not clear how consent can be obtained to use such personal data, or how to ensure that these data are stored safely and securely.



### A machine-learning algorithm to target COVID testing of travellers

Eva was developed in consultation with lawyers, who ensured that the program abided by the privacy protections afforded by the EU's General Data Protection Regulation (GDPR). Under the GDPR, organizations, such as airlines, that collect personal data need to follow security standards and obtain consent to store and use the data — and to share them with a public authority. The information collected tends to be restricted to the minimum amount required for the stated purpose.

But this is not necessarily the case outside the EU. Moreover, techniques such as machine learning that use AI are limited by the quality of the available data. Researchers have revealed many instances in which algorithms that were intended to improve decision-making in areas such as medicine and criminal justice reflect and perpetuate biases that are common in society. The field needs to develop standards to indicate when data — and the algorithms that learn from them — are of sufficient quality to be used to make important decisions in an emergency. There must also be a focus on transparency about how algorithms are designed and what data are used to train them.

The hunger with which Drakopoulos's offer of help was accepted shows how eager policymakers are to improve their ability to respond in an emergency. As such algorithms become increasingly prominent and more

widely accepted, it could be easy for them to slide, unnoticed, into day-to-day life, or be put to nefarious use. One example is that of facial-recognition technologies, which can be used to reduce criminal behaviour, but can also be abused to invade people's privacy (see [Nature 587, 354–358; 2020](#)). Although Eva's creators succeeded in doing what they set out to do, it's important to remember the limitations of big data and machine learning, and to develop ways to govern such techniques so that they can be quickly — and safely — deployed.

Despite a wealth of methods for collecting data, many policymakers have been unable to access and harness data during the pandemic. Researchers and funders should start laying the groundwork now for emergencies of the future, developing data-sharing agreements and privacy-protection protocols in advance to improve reaction times. And discussions should also begin about setting sensible limits on how much decision-making power an algorithm should be given in a crisis.

*Nature* **597**, 447-448 (2021)

doi: <https://doi.org/10.1038/d41586-021-02554-y>

---

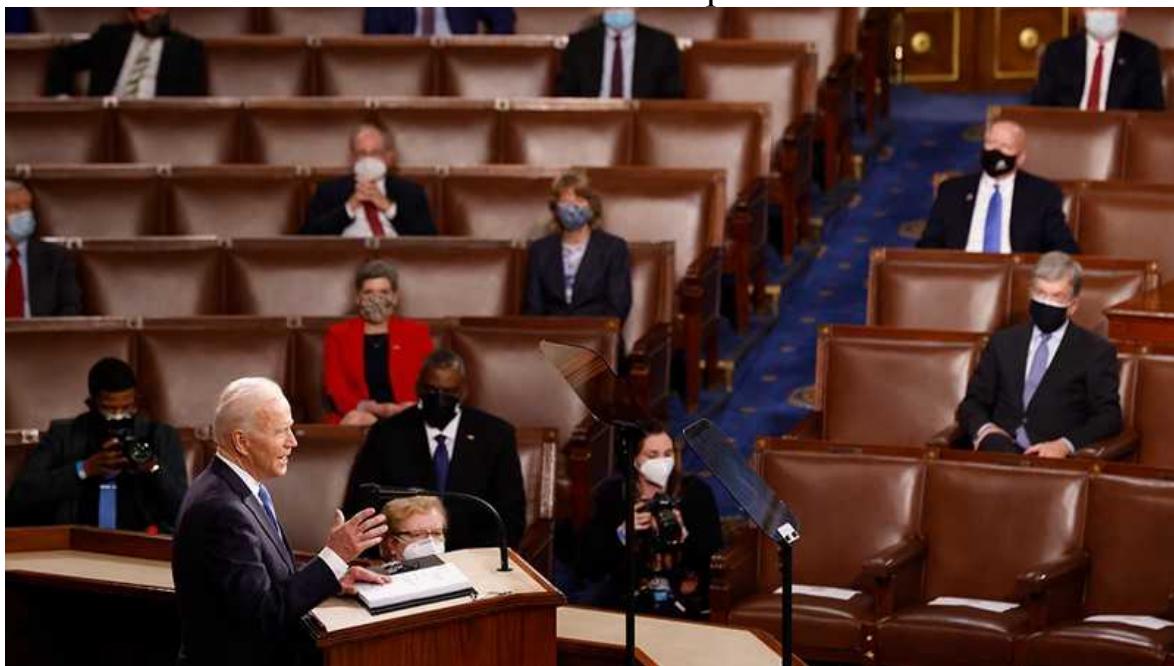
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02554-y>.

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

- EDITORIAL
- 21 September 2021

# Biden must keep funding pledge to historically Black colleges and universities

Congress has eviscerated the US president's pledge to upgrade research infrastructure at universities that serve underrepresented communities.



Joe Biden addressing a joint session of Congress in April. Democrat lawmakers must abide by his pledge to scale-up research in minority-serving institutions. Credit: Jonathan Ernst-Pool/Getty

It's a scandal: for decades, more than 400 colleges and universities in the United States that focus on educating students from under-represented communities, including Black, Hispanic and Indigenous Americans, have been underfunded — by both state and federal governments.

Efforts are now under way to address some of these historical injustices. Courts are awarding compensation to some of the more than 100 historically Black colleges and universities (HBCUs) that form part of a wider group known as minority-serving institutions (MSIs).

And in March this year, the White House proposed US\$20 billion to upgrade research infrastructure across MSIs as a whole. If approved by Congress, this funding would be dedicated to upgrading laboratories and creating new centres for research excellence — including a new national climate laboratory affiliated with an HBCU.

Fast forward six months, and the US Congress — which must approve the government's spending plans — has eviscerated US President Joe Biden's original proposal.



### [Springboard to science: the institutions that shaped Black researchers' careers](#)

On 8 September, Democrats in the House of Representatives introduced an education funding package that amounts to a fraction of the original \$20-billion request. Democrats are instead proposing just \$1.45 billion for MSIs from the federal government, to be distributed among the 400 institutions over 5 years between fiscal years 2022 and 2026.

It's a small increase from the roughly \$1 billion that the federal government annually spends on grants and scholarships at these universities. But it's nothing like what the Biden team acknowledged is needed to make up for decades of discrimination and neglect — by scaling up research across hundreds more higher-education institutions.

The House Democrats' proposal does include \$2 billion in federal grants earmarked for all US universities outside the R1 category (under the Carnegie Classification of Institutions of Higher Education), which indicates the highest levels of research activities. But more than 700 institutions will need to compete for this funding. "We are struck by the contrast between the vision laid out by the president and the actual application that we see in Congress," Lodriguez Murray, senior vice-president of public policy and government affairs at UNCF, an organization that raises funding for HBCUs, told *The Washington Post*.

HBCUs in the United States trace their origins to the segregation era of the 1800s. They seek to provide a nurturing environment for their students in a way that is less common elsewhere in higher education. The university experience is like being part of a family, several HBCU staff members and students [have told \*Nature\*](#). "It's not unusual for students who experience housing or food insecurity to be taken to an administrator's home and given care and support," said Ronald Smith, who runs mentoring programmes at Howard University in Washington DC.



## Discovering allyship at a historically Black university

The majority are teaching-focused institutions, although an increasing number have ambitions to excel at research, too. One-third of Black Americans with a PhD earned their first degree at an HBCU; 11 of these institutions are in the second-highest research classification, called R2, but none yet is among the 131 universities with the coveted R1 status.

For decades, HBCUs have suffered from under-investment — especially when compared with the funding of predominantly white institutions. Now, in addition to long-standing fundraising from UNCF, technology corporations are also stepping in with donations. Google is providing \$50 million to 10 HBCUs, and Apple \$5 million to four institutions.

Some HBCUs are also seeing extra funding from legal settlements in which state governments are compensating universities for past inequities. In the United States, state governments fund public universities and the federal government provides grants for research. Four HBCUs in Maryland — including Morgan State University in Baltimore — will share \$577 million from a settlement with the state of Maryland over the next decade, following a 15-year campaign by alumni highlighting that the state had treated its HBCUs less fairly than it did predominantly white institutions.

Such settlements are an overdue step, but the leaders of universities and colleges educating students from under-represented communities rightly say that there is no substitute for steady, predictable, long-term funding, as opposed to one-off grants — for which institutions that are intentionally collaborative and inclusive will have to start competing.

Institutions, agencies and governments around the world have made many pledges to increase inclusion in the past year in science, technology, engineering and mathematics. These pledges need to be fulfilled and words must now translate into action. That means congressional support for research at historically underfunded universities at a level that is much closer to the Biden administration's original \$20-billion proposal.

*doi: <https://doi.org/10.1038/d41586-021-02555-x>*

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02555-x>

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

- WORLD VIEW
- 21 September 2021

# Hurricane Ida shows the one-two punch of poverty and climate change



US President Joe Biden's environmental-justice adviser says: tackle inequality and global warming together.

- [Catherine Coleman Flowers](#) 0

I spent the 16th anniversary of Hurricane Katrina watching as the Weather Channel tracked Hurricane Ida. The two followed similar paths. As I write, the damage from Hurricane Ida is estimated at beyond US\$95 billion. Many in Louisiana still lack power, and more than 70 people are dead across 8 states.

Some of the worst-hit communities had it hard enough already. I know this from my work on *Waste* (The New Press, 2020) — exposing water contamination and poor sanitation in rural parts of the United States — for which I won a MacArthur Fellowship last year, and from my position as founding director of the Center for Rural Enterprise and Environmental Justice in Montgomery, Alabama. About two million people in the United States, including many around New Orleans, lack proper sewerage. Consequently, these regions have widespread hookworm infections, formerly thought to persist in only the poorest countries.

Advocates often talk about social justice, political justice, environmental justice, climate justice and more as though they are separate issues. The fact is: inequalities overlap and amplify each other. Those bearing the brunt of climate change often have the fewest resources and the most constraints on their civil rights, and live in the most polluted places. The pollution and warming that degrades farmlands and parklands disproportionately harms people of colour in cities. The communities that most need resilient, sustainable infrastructure can't afford it.



### [Five ways to ensure flood-risk research helps the most vulnerable](#)

St James Parish, a region of Louisiana along the Mississippi River that is flanked with dozens of petrochemical plants, exemplifies overlapping injustices. As other parts of the United States closed such plants, more

opened in and near the parish — known as Cancer Alley — and sickness rates rose. Earlier this year, the United Nations said the region exemplified environmental racism, citing US Environmental Protection Agency (EPA) data showing that cancer risks in predominantly Black areas in this parish are about 1.7 times those of predominantly white areas, and blamed systemic racism and a lack of federal regulations.

As I watched Hurricane Ida close in, I prayed for the people and homes I'd visited in the parishes of St James and St John the Baptist, where people who already have the country's highest cancer and COVID-19 rates now had to worry about losing their homes to storms made worse by climate change. Both parishes were pounded by Hurricane Ida and Hurricane Katrina. The plants there pollute the surroundings and contribute to climate change, and are inevitably placed near marginalized communities.

Recovery programmes launched after natural disasters rarely build resilience against a more dangerous climate, and they often further increase disparities. Research that tracked assistance from the US Federal Emergency Management Agency to people whose homes were damaged by hurricanes from 2005 to 2016 found that inspectors were less likely to visit areas with more Black residents, denying them a chance to apply for assistance. Even for damage inspections that were filed, those from Black homeowners were denied without reason almost three times as often as those from white homeowners. Separate work found that white residents of counties hit by a natural disaster saw their wealth grow, whereas that of Black residents in the same counties shrank.

The EPA has looked at projected climate-change impacts across the population. It has again found that the most severe harms from climate change fall disproportionately on underserved communities that are the least able to prepare for, and recover from, heat waves, poor air quality, flooding and other impacts.

Fighting such inequalities does not fall neatly under causes such as climate justice, social justice or environmental justice. Lately, I have been using the term planetary justice to encompass it all.



## How environmental racism is fuelling the coronavirus pandemic

I am heartened by US President Joe Biden's efforts to incorporate such thinking into policy. He has established the White House Environmental Justice Interagency Council, and the Justice40 Initiative, which requires that underprivileged communities accrue at least 40% of benefits from federal environmental investments. This means cleaning up legacy pollution, investing in clean energy, transportation and quality housing, and paying attention to those whose lives are the most precarious. Residents of frontline communities from St James and St John's parishes, to island nations, Indigenous peoples and developing nations, should have a seat at the table to work on solutions, which should be deployed to vulnerable communities first.

Both political will and funds are needed to make this happen. The UN has set up a mechanism it calls climate finance to funnel funds from robust nations to vulnerable ones, warning: "Without investing in the right places, the world will not achieve its climate goals." Even so, the International Renewable Energy Agency estimates that the world is underinvesting in clean-energy transitions by \$3 trillion annually.

But consider this: I am a Black woman from a rural community, one of the poorest regions of the United States, where concerns are more likely to be ignored than addressed. This year, I was invited to co-chair the first-ever

White House Environmental Justice Advisory Council, which Biden elevated from a little-known EPA committee. That is progress, and gives me hope for the future.

*Nature* **597**, 449 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-02520-8>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02520-8>

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

- RESEARCH HIGHLIGHT
- 15 September 2021

# A CRISPR fix for muscles hatches from a viral shell

Scientists create millions of mutant viruses to find those that excel at ferrying genome-editing tools into muscle tissue.

 Whole mount fluorescent images of quadriceps from 8-week-old C57BL/6J mice

Muscle from a mice injected with a standard viral shell (centre) takes up fewer viral particles than does muscle (right) injected with a viral shell called MyoAAV, as indicated by a fluorescent label. (The muscle on the left was injected with saline.) Credit: M. Tabebordbar *et al.*/*Cell*

Targeting genome-editing tools to the right part of the body could become much easier thanks to delivery vehicles made from a viral protein shell<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 450 (2021)

doi: <https://doi.org/10.1038/d41586-021-02508-4>

## **References**

1. 1.

Tabebordbar, M. *et al.* *Cell* **184**, 4919–4938 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02508-4>

- RESEARCH HIGHLIGHT
- 15 September 2021

## China faces one–two punches of extreme weather as Earth warms

The chances of heavy rain and high heat within one week are higher now than in the past two millennia, records suggest.



Floods that struck Japan in 2018 (pictured) were swiftly followed by extreme heat, a deadly duo that could also strike China in the coming decades. Credit: The Asahi Shimbun/Getty

China faces a growing risk of catastrophic flooding followed quickly by a lethal heatwave.

### Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 450 (2021)

doi: <https://doi.org/10.1038/d41586-021-02495-6>

## **References**

1. 1.

Liao, Z., Chen, Y., Li, W. & Zhai, P. *Geophys. Res. Lett.*  
<https://doi.org/10.1029/2021GL094505> (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02495-6>

- RESEARCH HIGHLIGHT
- 15 September 2021

# An unruly painkiller is tamed with inspiration from nature

Scientists borrow features of the body's receptor for tetrodotoxin to create a useful synthetic structure.

 Transmission electron micrograph of a section through a sciatic nerve.

The sciatic nerve (artificially coloured). A molecule that controls the release of the painkiller tetrodotoxin can have a long-lasting anaesthetic effect on the nerve. Credit: Steve Gschmeissner/Science Photo Library

Tetrodotoxin is a powerful painkiller, but a non-toxic form offering extended pain relief has proved elusive. Now researchers have succeeded in delivering slow-release tetrodotoxin safely by imitating the body's own receptor for the drug<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 450 (2021)

doi: <https://doi.org/10.1038/d41586-021-02514-6>

## **References**

1. 1.

Ji, T. *et al.* *Nature Biomed. Eng.* <https://doi.org/10.1038/s41551-021-00793-y> (2021).

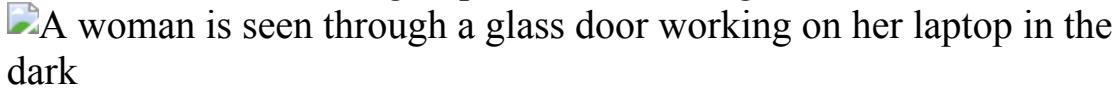
---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02514-6>

- RESEARCH HIGHLIGHT
- 14 September 2021
- Correction [15 September 2021](#)

# Home working brings longer hours, fewer phone calls

Data on more than 60,000 workers at Microsoft show that remote working led to more ‘siloeed’ work groups and the sending of more e-mails.



The pandemic-driven shift towards remote working led to an increase in working hours at one large US firm. Credit: Getty

The COVID-19 pandemic forced legions of workers to move from centralized offices to kitchen tables. An analysis of the activities of employees at one major technology firm suggests that widespread remote working can curb real-time communication and reduce collaboration between groups<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 450 (2021)

doi: <https://doi.org/10.1038/d41586-021-02496-5>

### **Updates & Corrections**

- **Correction 15 September 2021: Clarification 15 September 2021:**  
The headline of this article has been revised to clarify that even in the absence of a pandemic, a shift to remote working will change work patterns.

### **References**

1. 1.

Yang, L. *et al.* *Nature Hum. Behav.* <https://doi.org/10.1038/s41562-021-01196-4> (2021).

---

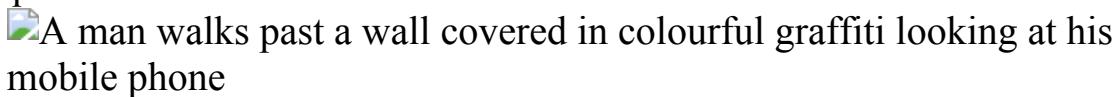
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02496-5>

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

- RESEARCH HIGHLIGHT
- 17 September 2021

# Humans walk efficiently even with their heads in the clouds

Exoskeleton-clad volunteers show that adapting to an energy-saving pace requires almost no attention.



Maintaining the most energy-efficient way of walking makes minimal cognitive demands, freeing up the brain to focus on other things, such as text messages. Credit: Julian Castle/Loop Images/UIG/Getty

Can you chew gum and walk? An analysis of gait and attention suggests that multitasking on a stroll should be a breeze<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 451 (2021)

doi: <https://doi.org/10.1038/d41586-021-02515-5>

## **References**

1. 1.

McAllister, M. J., Blair, R. L., Donelan, J. M. & Selinger, J. C. *J. Exp. Biol.* **224**, jeb242655 (2021).

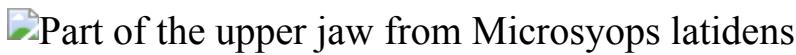
---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02515-5>

- RESEARCH HIGHLIGHT
- 14 September 2021

# A sweet tooth gave ancient primates a mouthful of woe

Fossils of a monkey-like animal that lived tens of millions of years ago furnish the earliest evidence of a mammal with cavities.



A fossilized jaw (computer reconstruction) of *Microsyops latidens* has cavities — seen as oval depressions — in the teeth second and third from the bottom. Credit: Keegan Selig

Even without fizzy drinks and sweets, an extinct primate had the same dental problems as modern humans, according to fossils — providing the earliest known evidence of mammals with cavities<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 451 (2021)

doi: <https://doi.org/10.1038/d41586-021-02475-w>

## **References**

1. 1.

Selig, K. R. & Silcox, M. T. *Sci. Rep.* **11**, 15920 (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02475-w>

- RESEARCH HIGHLIGHT
- 13 September 2021

# Puffins and friends suffer in washing-machine waves

Cyclones could make it difficult for seabirds such as little auks and puffins to hunt, which can lead to their starvation.



Cyclones churn the water so much that they make it hard for the Atlantic puffin to catch enough to eat. Credit: David Grémillet

After cyclones in the north Atlantic, droves of emaciated, dead seabirds can wash ashore on North American and European beaches. New research probes the cause of these mass-mortality events, called winter wrecks, and suggests that climate change might worsen the pattern<sup>1</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 451 (2021)

doi: <https://doi.org/10.1038/d41586-021-02494-7>

## **References**

1. 1.

Clairbaux, M. *et al. Curr. Biol.*  
<https://doi.org/10.1016/j.cub.2021.06.059> (2021)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02494-7>

# News in Focus

- **[Preprint ban reversal, vaccine boosters and awards bias](#)** [ 22 September 2021]  
News Round-Up • The latest science news, in brief.
- **[The fight to manufacture COVID vaccines in lower-income countries](#)** [ 15 September 2021]  
News • Drug companies and wealthy countries are facing increased pressure to partner with firms in the global south but are reluctant to relinquish control.
- **[Did the coronavirus jump from animals to people twice?](#)** [ 16 September 2021]  
News • A preliminary analysis of viral genomes suggests the COVID-19 pandemic might have multiple animal origins — but the findings still have to be peer reviewed.
- **[Australian bush fires belched out immense quantity of carbon](#)** [ 15 September 2021]  
News • The 2019–20 wildfires generated 700 million tonnes of carbon dioxide — but a lot of that might have been mopped up by phytoplankton in the ocean.
- **[New type of dark energy could solve Universe expansion mystery](#)** [ 17 September 2021]  
News • Hints of a previously unknown, primordial form of the substance could explain why the cosmos now seems to be expanding faster than theory predicts.
- **[Swedish research misconduct agency swamped with cases in first year](#)** [ 13 September 2021]  
News • The newly formed government organization tackled 46 research-fraud investigations in 2020 — three times as many as expected.
- **[How far will global population rise? Researchers can't agree](#)** [ 21 September 2021]  
News Feature • The United Nations forecasts that nearly 11 billion people will be living on Earth at the end of the century, but other demographic research groups project that population will peak earlier and at a much lower level.
- **[How dangerous is Africa's explosive Lake Kivu?](#)** [ 22 September 2021]

News Feature • An unusual lake in central Africa could one day release a vast cloud of greenhouse gases that suffocates millions of people. But it's not clear whether the threat is getting worse.

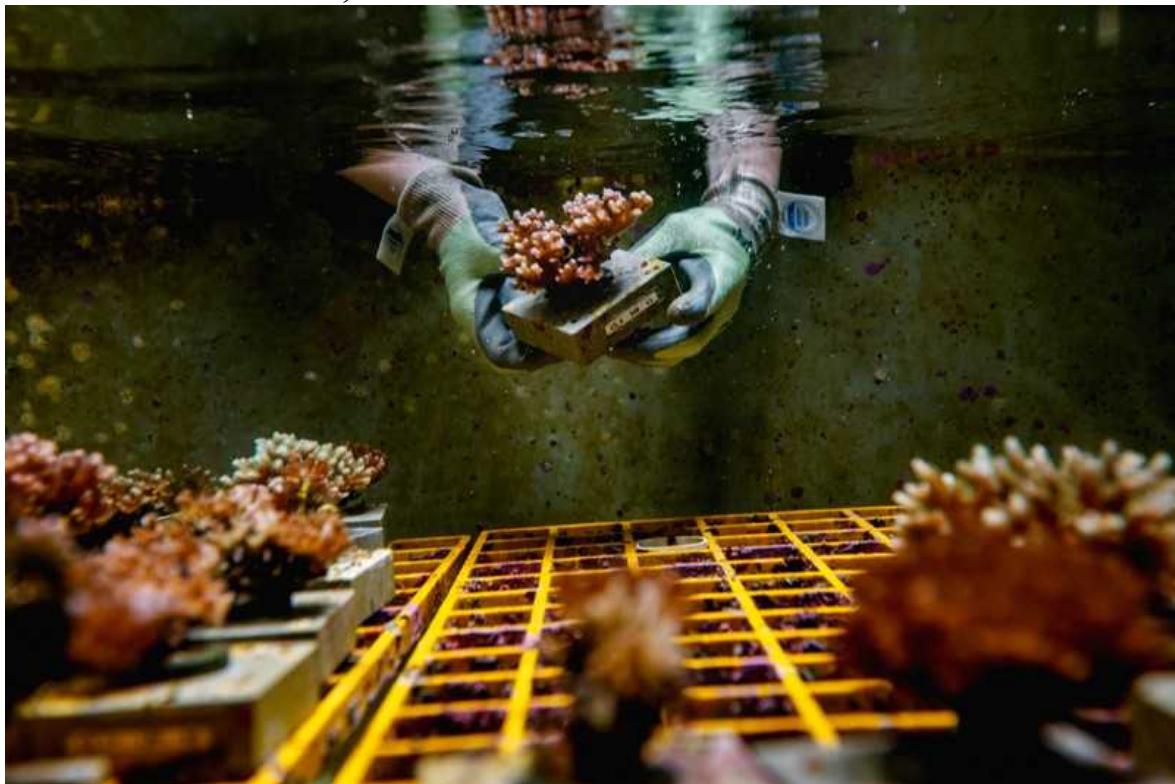
---

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

- NEWS ROUND-UP
- 22 September 2021

# Preprint ban reversal, vaccine boosters and awards bias

The latest science news, in brief.



Scientists study corals in a laboratory at the Australian Institute of Marine Science in Queensland. Credit: Jonas Gratzer/LightRocket/Getty

## Australian funder backflips on preprint ban

Australia's major research funding body has [backtracked on a rule that banned the mention of preprints in grant applications](#).

The policy adjustment by the Australian Research Council (ARC) comes four weeks after an anonymous researcher behind the ARC Tracker account on Twitter revealed that dozens of applications for early-career funding schemes had been rejected for citing preprints.

Announcing the U-turn in a statement on 14 September, the funder said the decision “reflects contemporary trends and the emerging significance of preprint acceptance and use across multiple research disciplines as a mechanism to expedite research”.

According to the statement, future grant applications will not be excluded for citing or including preprints — but the change will not apply automatically to applications that were ruled ineligible or are currently under review.

Some researchers welcome the reversal, but others say that the move does not go far enough. Nick Enfield, an anthropologist at the University of Sydney, applauds the decision but finds it “regrettable that eligibility rulings haven’t been overturned”.



A resident of an assisted-living facility in Israel receives a third dose of COVID-19 vaccine.Credit: Nir Alon/Zuma Press

## COVID vaccine booster shows promise

Older Israelis who have received a third dose of a COVID-19 vaccine are [much less likely to test positive for SARS-CoV-2 or to develop severe COVID-19](#) than are those who have had only two jabs, according to a highly anticipated study published on 15 September ([Y. M. Bar-On et al. N. Engl. J. Med. https://doi.org/gmtzb3; 2021](#)).

The study evaluated 1.1 million Israelis over the age of 60 who had received their first two doses at least five months earlier. Twelve or more days after receiving a third jab, participants were about 19.5 times less likely to have severe COVID-19 than were people in the same age group who had received only two jabs and were studied during a similar time period.

“It’s a very strong result,” says Susan Ellenberg, a biostatistician at the University of Pennsylvania in Philadelphia, who adds that the data might be the most robust she has seen in favour of boosters. But potential biases in the data leave some scientists unconvinced that boosters are necessary for all populations — and the data do not dispel concerns about vaccine equity when billions of people are still waiting for their first jab.



Winners at the L’Oreal-UNESCO Awards For Women in Science International in 2016. L–R: Jennifer Doudna, Hualan Chen, Andrea Gamarnik, Quarraisha Abdool Karim and Emmanuelle Charpentier.Credit: Bertrand Rindoff Petroff/Fondation L’Oreal/Getty

## Women less likely to win major research awards

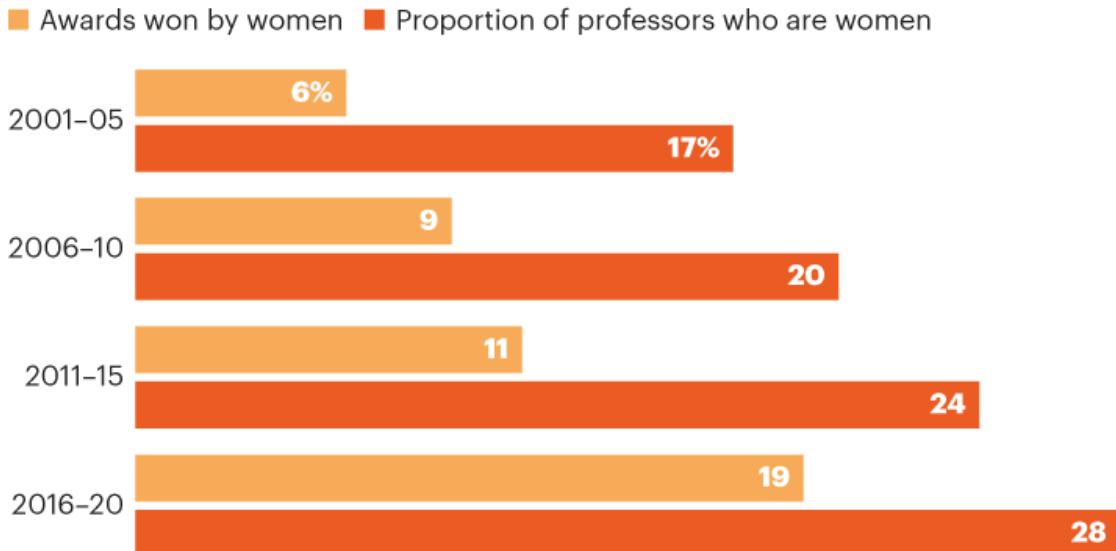
Women’s share of prizes rewarding research excellence is increasing, but still [lags behind the proportion of professorial positions held by women](#), according to an analysis of 141 leading science prizes awarded over the past 2 decades.

Lokman Meho, an information scientist at the American University of Beirut, examined whether gains in professorships for women have translated into awards honouring their work ([L. I. Meho \*Quant. Sci. Stud.\* <https://doi.org/gwdn>; 2021](https://doi.org/gwdn)).

His findings show a narrowing but persistent gender gap in the highest tiers of awards (see ‘Closing the gap’).

## CLOSING THE GAP

Although the proportion of professors who are women has increased in recent decades, the percentage of top research prizes going to women still lags behind.



©nature

Source: Meho, L. I. *Quant. Sci. Stud.* [https://doi.org/10.1162/qss\\_a\\_00148](https://doi.org/10.1162/qss_a_00148) (2021).

Hans Petter Graver, a legal scholar and president of the Norwegian Academy of Science and Letters in Oslo, which administers the Abel and Kavli prizes, says the results send “a signal to institutions awarding prestigious science prizes to do more for diversity”.

Women have comparable publication and citation rates to men, but tend to have shorter careers and publish fewer papers as first or last author, according to other studies.

In his analysis, Meho identified 141 highly prestigious international prizes — including the Nobel prizes, the Fields Medal for mathematics and the Robert Koch Award for biomedical sciences — awarded to 2,011 men and 262 women between 2001 and 2020. He grouped recipients into five-year intervals.

The results show that the number of awards honouring female scientists has increased in the past 20 years, but women still remain under-represented. “We are moving in the right direction, although slowly,” Meho says.

Although the study did not examine causes of gender bias, he argues that women are not receiving fewer awards because of the quality or quantity of their research.

Instead, he puts it down to implicit bias, coupled with a lack of proactive efforts to address inequities in science.

Around two-thirds of the 141 awards recognized women at some point between 2016 and 2020, up from 30% of the 111 awards that were offered in 2001–05. Women’s average share of the prizes, when counting all recipients, neared 20% in 2016–20.

However, this fell short of the proportion of professorial positions held by women over the same period, Meho found.

*Nature* **597**, 453 (2021)

doi: <https://doi.org/10.1038/d41586-021-02521-7>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02521-7>

- NEWS
- 15 September 2021
- Correction [16 September 2021](#)

# The fight to manufacture COVID vaccines in lower-income countries

Drug companies and wealthy countries are facing increased pressure to partner with firms in the global south but are reluctant to relinquish control.

- [Amy Maxmen](#)



In Mumbai, India, people stand in queue to receive the COVID-19 vaccine. Credit: Divyakant Solanki/EPA-EFE/Shutterstock

Vaccines against COVID-19 are not reaching many people in the global south, despite donations from wealthy nations. Less than 1% of people in low-income countries are fully vaccinated, and just 10% are in lower-middle-income countries, compared with more than half in high-income countries.

Many researchers say the best way to ensure equitable access to COVID-19 vaccines is to enable countries in the global south to make their own.

“Charity is good, but we can’t rely on charity alone,” says Peter Singer, an adviser to the director-general of the World Health Organization (WHO).

Since last year, health-advocacy organizations have been pressing pharmaceutical companies and governments that developed highly effective vaccines to share their patented knowledge and technology with drug manufacturers that could produce them for poorer countries. These vaccines include the messenger-RNA jabs created by Moderna in Cambridge, Massachusetts, and Pfizer in New York City and BioNTech in Mainz, Germany, and a viral-vector vaccine developed by Johnson & Johnson (J&J) in New Brunswick, New Jersey.



[Why a pioneering plan to distribute COVID vaccines equitably must succeed](#)

Calls to manufacture more vaccines in the global south have grown louder in advance of high-level pandemic discussions at the United Nations General Assembly, which began this week, and a US-led, Global COVID-19 Summit on 22 September. Advocates are clamouring for a variety of approaches. Some had pointed to the deployment of the Sputnik V vaccine as a model of pandemic diplomacy. Russia broadly licensed the jab to 34 drug companies outside its borders, including several in India and Brazil. But manufacturers are now saying that the second dose of the vaccine — which has a different composition than the first — is difficult to produce in large quantities.

In a [letter signed by several Indian civil society groups](#) — shared with *Nature* — advocates are urging US President Joe Biden to compel J&J to partner with drug companies in the global south, arguing that those making Sputnik V could easily pivot to the J&J vaccine because they rely on similar technologies. They estimate that the transition would take less than six months.

Achal Prabhala, an author on the letter and a coordinator at AccessIBSA, a medicines-access initiative in Bengaluru, India, thinks this switch would help to quickly protect people in places lacking vaccines (see ‘Protection divide’). He adds that partnerships with the companies that developed mRNA vaccines will also be crucial because of the shots’ effectiveness and adaptability. India, in particular, could help to tame the pandemic if the country was enabled to make more shots, he says, illustrated by its role in providing the majority of vaccines against other diseases to low- and lower-middle-income countries. “For 3.9 billion people, we are the bulwark of vaccine manufacturing. So, if there aren’t contracts here, the world suffers.”

## PROTECTION DIVIDE

So far, 55% of the people living in high-income countries have been fully vaccinated against COVID-19, but less than 1% of the residents of low-income countries have been fully vaccinated.



\*Data are as of 8 September

©nature

Credit: KFF and Our World in Data

Such calls have not yet gained traction. Outside of deals to bottle and package their vaccines, J&J has only one partnership with an Indian company, and Pfizer, BioNTech and Moderna have none in India, South America or Africa. Pharmaceutical companies have cited reasons including quality concerns and the time required to get new companies up to speed. Instead, they say they're ramping up their own production, and they ask wealthy nations to increase vaccine donations to poorer ones. Prabhala calls their arguments "a useful canard that obscures the real barrier — an unwillingness on the part of western pharmaceutical companies to relinquish control over their patents and technology, even at the cost of millions of lives".

Although the Biden administration supported a waiver on intellectual property surrounding COVID-19 vaccines that was proposed by India and South Africa at a World Trade Organization meeting last October, action has stalled. And the administration has not pushed US companies to partner with those in the global south. Germany, which funded the development of BioNTech's mRNA vaccine, later licensed to Pfizer, remains opposed to patent waivers.

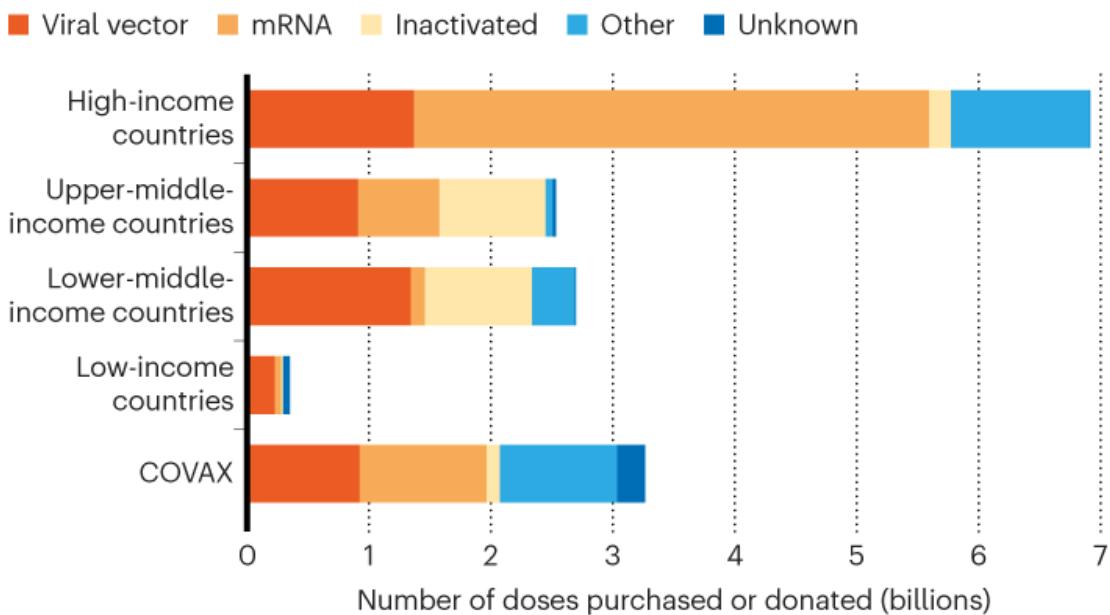
As months pass, some researchers have stopped hoping for partnerships to come to fruition. A group in South Africa has decided to try and re-create existing vaccines. Others argue that funds would be best spent on getting manufacturers in the global south prepared to pump out the next generation of vaccines currently in clinical trials. Most global health researchers agree that regional manufacturing is the only way to ensure worldwide vaccination in a crisis. Shahid Jameel, a virologist at the Trivedi School of Biosciences at Ashoka University in New Delhi, says, “We can’t fix vaccine inequalities until vaccine manufacturing is distributed.”

## Low yields

Companies might produce an estimated 12 billion doses of COVID-19 vaccines this year, but many more are needed, says Andrea Taylor, a global health researcher who leads a [vaccine-tracking project](#) at Duke University in Durham, North Carolina. Many wealthy nations have purchased enough doses to cover their populations several times over while some countries have very few, she says. The type of vaccine in demand has shifted, too. China’s vaccines, made from inactivated SARS-CoV-2 coronaviruses, accounted for nearly a third of jabs in lower-income countries through August. But questions about the shots’ efficacy have some countries searching for other options. Meanwhile, demand for mRNA vaccines has soared because wealthy countries are recommending [third doses](#) to, in theory, boost their populations’ immunity (see ‘Dose distribution’).

## DOSE DISTRIBUTION

Data on how many vaccines countries have acquired are poorly reported. But researchers can track doses purchased by, or donated to, countries — understanding that many of these doses have yet to be delivered. By late August, for example, the international alliance COVAX had delivered only about 230 million doses. High-income countries have purchased the majority of mRNA vaccines.



©nature

Credit: Duke Global Health Innovation Center

Lacking mRNA options, many nations in the global south rely on viral-vector shots that use a harmless inactivated virus to deliver their payload to cells. Indeed, 88% of the people vaccinated in India have gotten viral-vector shots developed by the University of Oxford and AstraZeneca in the United Kingdom — and produced by the Serum Institute of India, the biggest vaccine manufacturer in the world. International organizations leading COVID-19 Vaccines Global Access (COVAX), a system to supply COVID-19 vaccines to low- and middle-income countries, expected the Serum Institute to provide a bulk of their vaccines, but that plan fell short when the Indian government restricted exports in March when the country faced a deadly surge of COVID-19 and only 2% of its population had been vaccinated. Because of issues including the export pause and a lack of

donations, COVAX has shifted its goal of delivering two billion doses from this year to 2022.



The Serum Institute of India, the world's largest vaccine manufacturer was barred from exporting its version of the AstraZeneca COVID-19 vaccine. Credit: Dhiraj Singh/Bloomberg/Getty

Russia's Sputnik V vaccine can't bolster COVAX's supply because it isn't authorized by the WHO, despite its authorization in India, Brazil and dozens of other countries. The organization has given the green light to J&J's jab, however — another reason that advocates support a transition to that shot. Handing off Sputnik V wasn't simple, but manufacturers say the technology transfer process is instructive. Russian scientists gave willing drug companies essential ingredients for the vaccine and lists of equipment and supplies, and they visited the plants to teach them the manufacturing process.

Hemanth Nandigala, the managing director of one company producing Sputnik V, Virchow Biotech in Hyderabad, India, says that such "hand holding" made the technology transfer faster — about three months —

although scaling production, passing regulatory clearances and commercialization has taken another five months. Only in September had it become clear that many companies making Sputnik V have low yields for the second dose. The vector in the first shot — adenovirus 26 — is similar to that in J&J’s vaccine, so Nandigala says that, if enabled, companies could reorient their processes to produce this jab.

J&J did not respond to requests from *Nature* about why it has not partnered with more companies in the global south. However, at a [7 September press briefing](#), J&J’s chief scientific officer, Paul Stoffels, explained that transferring technology requires time to train a workforce to produce new and complex products. The company has partnered with one Indian company based in Hyderabad: Biological E. Its managing director, Mahima Datla, says that the transfer and scale-up took around seven months, and they hope to soon produce more than 40 million doses monthly. It’s not clear how many of those J&J vaccines will serve lower-income countries. “The decision on where they will be exported, and at what price, is under the purview of J&J completely,” she says.

In South Africa, J&J’s partnership with a pharmaceutical company in Durban that bottles its vaccines caused controversy after [The New York Times reported](#) that it was shipping the shots to Europe, despite more than 90% of people in South Africa having received no vaccine at all. Following the outcry, J&J said that future doses produced in South Africa would stay in Africa.

Faced with J&J’s reluctance, the authors of the letter from India argue that the US government gave the company \$1 billion to develop their technology, and could therefore compel it to increase its output by partnering with the 34 firms outfitted for Sputnik V. “If US President Biden is indeed serious about vaccinating the world, his administration has the moral, legal, and if necessary, financial power to lift intellectual property barriers and persuade J&J to license its vaccine, with technology and assistance included, to every manufacturer currently engaged in making the Sputnik V vaccine,” they write.

## **Unsatisfied with leftovers**

When it comes to mRNA vaccines, researchers say transferring the knowledge and acquiring tools required for manufacturing will be challenging because of the newness of the technology. Nonetheless, Aditya Kumar, a representative for India's Stelis Biopharma in Bengaluru, which is producing Sputnik V, says that the steep learning curve would be worthwhile because mRNA vaccines seem to be simpler to make in large quantities than those based on viral vectors — a finicky process that requires researchers to grow adenoviruses in living mammalian cells. "Manufacturers like us are always considering how to scale vaccines because we understand the massive needs of the underserved world," he says.



### What it will take to vaccinate the world against COVID-19

For several months, the WHO has called on companies to share their licenses. At a press briefing last week, WHO director-general Tedros Adhanom Ghebreyesus said, "I will not stay silent when the companies and countries that control the global supply of vaccines think the world's poor should be satisfied with leftovers."

Specifically, Soumya Swaminathan, the WHO's chief scientist, has asked innovating firms to contribute their intellectual property to the Medicines Patent Pool, a United Nations-backed organization that aims to bring inexpensive drugs to poor countries. The group helps companies to forge partnerships by identifying reliable manufacturers, assisting with regulatory

approvals and finding licensing arrangements that offer vaccine developers royalties on drugs sold.

But the pharmaceutical industry hasn't played ball. Thomas Cueni, the director-general at the International Federation of Pharmaceutical Manufacturers & Associations (IFPMA), says the surest way to scale up manufacturing is to do it in-house. "It comes down to quality control and quality assurance, which is incredibly complex," he says, wherever you are in the world. "People can talk about additional partnerships, but they underestimate the challenge."

The chair of the patent pool, Marie-Paule Kieny, disagrees, pointing out that many researchers in her group previously worked at leading pharmaceutical firms, and have experience ensuring best practices. "The Medicines Patent Pool does not give licenses to manufacturers working in a garage," she says.



### [COVID vaccines to reach poorest countries in 2023 — despite recent pledges](#)

Another approach, Swaminathan says, is for companies to participate in a technology transfer hub in South Africa — announced by the WHO in June — where researchers who developed mRNA vaccines can teach other manufacturers how to make them. But Pfizer's chief executive officer, Albert Bourla, knocked the initiative at an IFPMA [press briefing last week](#), suggesting it would take "years" for companies to get up to speed. He added that vaccine supply won't be a problem next year, once Pfizer and other

firms have ramped up manufacturing. In an e-mail to *Nature*, a spokesperson for Pfizer explained that the company has initially relied on manufacturers in Europe and the United States to safely ramp up production, but might bring on more manufacturers in the future so that it can make up to four billion doses in 2022. Moderna did not respond to requests for comment from *Nature*. But its chief executive, Stéphane Bancel, [told analysts in May](#) that he was strongly opposed to patent waivers, and that outside companies would take upwards of 12 to 18 months to produce Moderna's mRNA shot.

Suhaib Siddiqi, a former director of chemistry at Moderna, based in Boston, Massachusetts, contests these long timelines. He argues that Moderna tested the efficacy of its vaccine and scaled up production within nine months, and therefore could teach experienced drug companies in India how to do the same. The WHO may share Siddiqi's confidence: Yesterday, [Reuters](#) broke the news that the South African hub will attempt to re-create Moderna's shot. Swaminathan confirms the report, adding that researchers familiar with the process have offered to assist.

An advocacy group based in Washington DC, Public Citizen, [argues](#) that the US Department of Health and Human Services (HHS) could help the WHO, too. They assert that HHS has rights over such information because the government invested US\$1.4 billion in the vaccine's development in 2020, in exchange for "access to all documentation and data" generated under a [publicly available contract](#) with Moderna. HHS declined to comment to *Nature* on the terms of the contract and Public Citizen's request.

## An easier path?

Instead of holding out for today's popular vaccines, some researchers hope that those in clinical trials will be easier to license and make in the global south. At the top of the list are protein-subunit vaccines, in which peptides matching those from SARS-CoV-2 teach the immune system to recognize the virus and fight it off. Researchers say the benefit of such vaccines is that vats of yeast or insect cells can churn out huge quantities of peptides, making the vaccines scalable. They add that many companies are familiar with the process because they produce vaccines for other diseases and recombinant drugs in a similar fashion.



### COVID boosters for wealthy nations spark outrage

One of these is a product from Biological E, which licensed the technology from Baylor College of Medicine and Texas Children's Hospital in Houston, Texas. Biological E's vaccine is currently in phase 3 clinical trials in India, and Datla expects it to be authorized by the Indian government by November and the WHO in January. Prashant Yadav, a health-care supply-chain specialist at the Center for Global Development in Washington DC, argues that it's worth waiting to see whether this type of vaccine is effective, because transferring the technology and scaling up production might be simpler than moving the needle with other companies. "If the protein-subunit vaccines work well, I'd put my money there," says Yadav.

But how a fresh set of companies will license their vaccines to external manufacturers remains to be seen. Peter Hotez, a vaccine researcher who helped to develop the subunit vaccine at Baylor, says they aren't putting patent restrictions on the technology, so that manufacturers in India, Indonesia and elsewhere can make billions of doses next year for the developing world. He says, "We're receiving calls weekly from low- and middle-income countries desperate for our vaccine."

*Nature* **597**, 455-457 (2021)

doi: <https://doi.org/10.1038/d41586-021-02383-z>

## Updates & Corrections

- **Correction 16 September 2021:** An earlier version of this article misquoted Thomas Cueni as referring to a list of 60 to 70 manufacturers in developing countries. In fact, he was referring to a checklist of 60 to 70 quality control criteria for manufacturers in developing countries, and he had noted that quality control is complex everywhere in the world. The story also incorrectly stated that the Serum Institute is contributing vaccines to COVAX. Although COVAX is expecting doses, an anonymous source with the Indian government [told the Hindustan Times](#) that the country's vaccine export ban stands.

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02383-z>

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

- NEWS
- 16 September 2021

# Did the coronavirus jump from animals to people twice?

A preliminary analysis of viral genomes suggests the COVID-19 pandemic might have multiple animal origins — but the findings still have to be peer reviewed.

- [Smriti Mallapaty](#)



Raccoon dogs — pictured here at a fur farm in China's Hebei province — are susceptible to SARS-CoV-2 and were sold at multiple markets in Wuhan. Credit: Greg Baker/AFP via Getty

SARS-CoV-2, the virus that causes COVID-19, could have spilled from animals to people multiple times, according to a preliminary analysis of viral genomes sampled from people infected in China and elsewhere early in the pandemic.

If confirmed by further analyses, the findings would add weight to the hypothesis that the pandemic originated in multiple markets in Wuhan, China, and make the hypothesis that SARS-CoV-2 escaped from a laboratory less likely, say some researchers. But the data need to be verified, and the analysis has not yet been peer reviewed.



#### [WHO report into COVID pandemic origins zeroes in on animal markets, not labs](#)

The earliest viral sequences, taken from people infected in late 2019 and early 2020, are split into two broad lineages, known as A and B, which have key genetic differences.

Lineage B has become the dominant lineage globally and includes samples taken from people who visited the Huanan seafood market in Wuhan, which also sold wild animals. Lineage A spread within China, and includes samples from people linked to other markets in Wuhan.

A crucial question is how the two viral lineages are related. If viruses in lineage A evolved from those in lineage B, or vice versa, that would suggest

that the progenitor of the virus jumped just once from animals to people. But if the two lineages have separate origins, then there might have been multiple spillover events.

## Dagger in the heart

The latest analysis — posted on the [virological.org discussion forum](https://virological.org/discussion/forum) — adds weight to the second possibility by questioning the existence of genomes linking the lineages.

The finding could be the “dagger into the heart” of the hypothesis that SARS-CoV-2 escaped from a lab, rather than originating from the wildlife trade, says Robert Garry, a virologist at Tulane University in New Orleans, Louisiana. But others say that more research is needed, especially given the limited genomic data from early in the pandemic.

“It is a very significant study,” says Garry. “If you can show that A and B are two separate lineages and there were two spillovers, it all but eliminates the idea that it came from a lab.”



### [The COVID lab-leak hypothesis: what scientists do and don't know](#)

The findings are “consistent with there being at least two introductions of SARS-CoV-2 into the human population”, says David Robertson, a

virologist at the University of Glasgow, UK.

Lineages A and B are defined by two key nucleotide differences. But some of the earliest genomes have a combination of these differences. Researchers previously thought that these genomes could be those of viruses at intermediate stages of evolution linking the two lineages.

But the researchers behind the latest analysis looked at them in detail and noticed some problems.

## Fine-tooth comb

They analysed 1,716 SARS-CoV-2 genomes in a popular online genome repository called GISAID that were collected before 28 February 2020, and identified 38 such ‘intermediate’ genomes.

But when they looked at the sequences more closely, they found that many of these also contained mutations in other regions of their genomes. And they say that these mutations are definitively associated with either lineage A or lineage B — which discredits the idea that the corresponding viral genomes date to an intermediate stage of evolution between the two lineages.

The authors suggest that a laboratory or computer error probably occurred in sequencing one of the two mutations in these ‘intermediate’ genomes. “The more we dug, the more it looked like, maybe we can’t trust any of the ‘transitional’ genomes,” says study co-author Michael Worobey, an evolutionary biologist at the University of Arizona in Tucson.

Such sequencing errors are not unusual, say researchers. Software can sometimes fill in gaps in the raw data with incorrect sequences, and viral samples can become contaminated, notes Richard Neher, a computational biologist at the University of Basel in Switzerland. “Such mishaps are not surprising,” he says. “Especially early in the pandemic, when protocols weren’t very established and people tried to generate data as fast as they could.”



### [After the WHO report: what's next in the search for COVID's origins](#)

Several researchers who sequenced samples included in the study told *Nature* it is unlikely that their sequences include errors in the two key nucleotides.

But the study authors counter that even if some of the genomes were sequenced correctly, other parts of the same genomes, or the locations from which the samples were collected, still clearly indicate that they belong to only one or the other lineage.

“It is very unlikely” that any of the ‘intermediate’ genomes are actually transitional genomes, says study co-author Joel Wertheim, a molecular epidemiologist at the University of California, San Diego.

Xiaowei Jiang, an evolutionary biologist at Xi'an Jiaotong–Liverpool University in Suzhou, China, says that the team behind the study must verify the findings by getting “the original raw sequencing data for as many genomes as possible”.

## Many markets

If the virus did jump between animals and people on several occasions, the fact that lineages A and B are linked to people who visited different markets in Wuhan suggests that multiple individual animals, of one or more species,

that were carrying a progenitor of SARS-CoV-2 could have been transported across Wuhan, infecting people in at least two locations.

A study published in June<sup>1</sup> found that live animals susceptible to SARS-CoV-2, such as raccoon dogs and mink, were sold in numerous markets in Wuhan. Previous studies<sup>2</sup> of the virus that caused severe acute respiratory syndrome (SARS) have concluded that it, too, probably jumped multiple times from animals to people.

The latest study, if verified, would mean that the scenario of a researcher accidentally being infected in a lab, and then spreading the virus to the population at large, would have had to happen twice, says Garry. It's much more likely that the pandemic had its origins in the wildlife trade, he says.

To gather more evidence, the team behind the latest analysis now plans to run computer simulations to test how well multiple spillovers would fit with the diversity of known SARS-CoV-2 genomes.

*Nature* **597**, 458-459 (2021)

doi: <https://doi.org/10.1038/d41586-021-02519-1>

## References

1. 1.

Xiao, X. *et al. Sci. Rep.* **11**, 11898 (2021).

2. 2.

Wang, L. F. & Eaton, B. T. *Curr. Top. Microbiol. Immunol.* **315**, 325–344 (2007).

- NEWS
- 15 September 2021

# Australian bush fires belched out immense quantity of carbon

The 2019–20 wildfires generated 700 million tonnes of carbon dioxide — but a lot of that might have been mopped up by phytoplankton in the ocean.

- [Smriti Mallapaty](#)



Sydney is encircled by huge bush fires that shroud it in smoke on 21 December 2019.Credit: Orbital Horizon/Copernicus Sentinel Data 2020/Gallo Images/Getty

The extreme bush fires that blazed across southeastern Australia in late 2019 and early 2020 released 715 million tonnes of carbon dioxide into the air — more than double the emissions previously estimated from satellite data, according to an analysis<sup>1</sup> published today in *Nature*.

“That is a stupendous amount,” says David Bowman, a fire ecologist at the University of Tasmania in Hobart, who adds that scientists might have to rethink the impact on global climate of extreme blazes, which have now raged not just across Australia, but across the [western United States](#) and Siberia. “Fire is a really big deal now.”

It’s not all bad news, however. Another paper<sup>2</sup> in *Nature* reports that much of this plume of carbon might have been indirectly sucked up by a gigantic phytoplankton bloom in the Southern Ocean.

## Worst fires on record

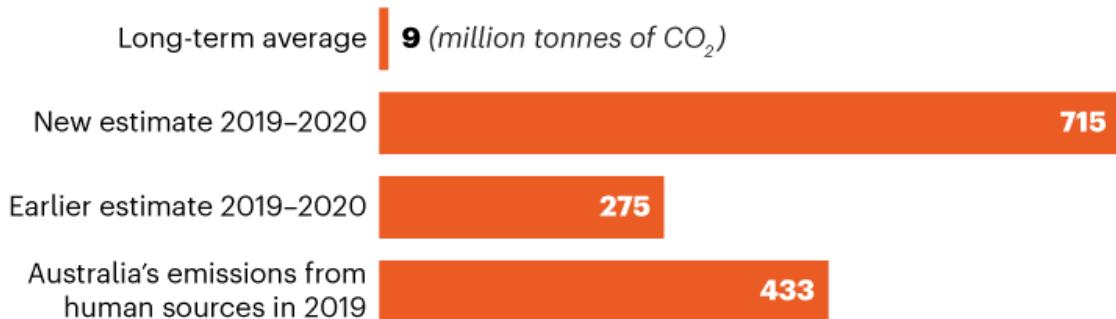
The [unprecedented fires](#) burnt across as much as 74,000 square kilometres of mostly eucalyptus, or gum, forest in southeast Australia — an area larger than Sri Lanka.

Previous estimates from global databases of wildfire emissions based on satellite data suggested that the fires released about 275 million tonnes of carbon dioxide during their zenith, between November 2019 and January 2020.

But the new analysis indicates that this figure was a gross underestimate, says Ivar van der Velde, lead author of the first paper. “These models often lack the spatio-temporal detail to explain the full impact these fires have,” says van der Velde, an environmental scientist at the SRON Netherlands Institute for Space Research, in Utrecht, and at the Free University of Amsterdam.

## RECORD EMISSIONS

Devastating fires in southeastern Australia in the summer of 2019–2020 released almost 80 times as much carbon dioxide into the atmosphere as a typical summer bush-fire season.



©nature

Source: Ref. 1/M. Crippa *et al.* ‘[Fossil CO<sub>2</sub> emissions of all world countries: 2020 report](#)’ (European Commission Joint Research Centre, 2020).

He and his team set out to get a better estimate, based on more-granular data from the tropospheric monitoring instrument TROPOMI on the European Space Agency’s Sentinel-5 Precursor satellite.

TROPOMI takes daily snapshots of carbon monoxide levels in the atmospheric column beneath it. The researchers used this data to calculate a more accurate estimate of the carbon monoxide emissions from the bush fires, which they used as a proxy for calculating carbon dioxide emissions.

Their final figure — 715 million tonnes — is nearly 80 times the typical amount of carbon dioxide emitted from fires in southeast Australia during the three peak months of the summer bush-fire season (see ‘Record emissions’).



## Enormous wildfires spark scramble to improve fire models

Bowman says the figure is similar to what his team calculated from the area of forests burnt<sup>3</sup>, but much higher than figures based on previous satellite measurements of emissions.

The key question is how these forests will recover, says Cristina Santín, a wildfire researcher at the Spanish National Research Council in Asturias. Wildfires have long been considered net-zero-carbon events, because the emissions they release are recaptured when the vegetation regrows — but an increase in the frequency and intensity of fires in Australia could mean that ecosystems never fully bounce back. If these fires “threaten the recovery of the ecosystem, then we really need to worry”, she says.

## **Reason to hope**

The second paper, also published today, could give researchers reason to hope, however. It suggests that the emissions generated by the bush-fire crisis were nearly offset by gigantic phytoplankton blooms in the Southern Ocean, recorded over the summer of 2019–20.

The findings demonstrate how wildfires can directly influence ocean processes, says study co-author Richard Matear, a climate scientist based in

Hobart with the Australian government's Commonwealth Scientific and Industrial Research Organisation. "The systems are connected."



### Climate change made Australia's devastating fire season 30% more likely.

He and his colleagues found that, during the fires, vast black plumes of smoke, rich in nutrients, were swept thousands of kilometres away over the ocean. Within days, these aerosols had infused the waters with much-needed iron, nourishing phytoplankton, which sucked up carbon equivalent to as much as 95% of the emissions from the fires.

The ocean seems to achieve "an amazing sleight of hand — like a magician", says Bowman. But he and other researchers say that more work needs to be done to understand where the carbon taken up by the plankton eventually goes, and whether it makes it back out into the atmosphere.

Both studies reveal surprising findings showing that "we don't understand fires as much as we really need to", says Santín — something she says we need to get a better handle on, because "fires are going to be increasingly important in the carbon cycle".

*Nature* **597**, 459-460 (2021)

doi: <https://doi.org/10.1038/d41586-021-02509-3>

# References

1. 1.

van der Velde, I. R. *et al.* *Nature* <https://doi.org/10.1038/s41586-021-03712-y> (2021).

2. 2.

Tang, W. *et al.* *Nature* <https://doi.org/10.1038/s41586-021-03805-8> (2021).

3. 3.

Bowman, D. M. J. S., Williamson, G. J., Price, O. F., Ndalila, M. N. & Bradstock, R. A. *Plant Cell Environ.* <https://doi.org/10.1111/pce.13916> (2020).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02509-3>

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

- NEWS
- 17 September 2021

# New type of dark energy could solve Universe expansion mystery

Hints of a previously unknown, primordial form of the substance could explain why the cosmos now seems to be expanding faster than theory predicts.

- [Davide Castelvecchi](#)



Data from the Atacama Cosmology Telescope suggest the existence of two types of dark energy at the very start of the Universe. Credit: Giulio Ercolani/Alamy

Cosmologists have found signs that a second type of dark energy — the ubiquitous but enigmatic substance that is pushing the current Universe's expansion to accelerate — might have existed in the first 300,000 years after the Big Bang.

Two separate studies — both posted on the arXiv preprint server in the past week<sup>1,2</sup> — have detected a tentative first trace of this ‘early dark energy’ in data collected between 2013 and 2016 by the [Atacama Cosmology Telescope](#) (ACT) in Chile. If the findings are confirmed, they could help to solve a long-standing conundrum surrounding data about the early Universe, which seem to be [incompatible with the rate of cosmic expansion](#) measured today. But the data are preliminary and don’t show definitively whether this form of dark energy really existed.

“There are a number of reasons to be careful to take this as a discovery of new physics,” says Silvia Galli, a cosmologist at the Paris Institute of Astrophysics.

The authors of both preprints — one posted by the ACT team, and the other by an independent group — admit that the data are not yet strong enough to detect early dark energy with high confidence. But they say that further observations from the ACT and another observatory, the South Pole Telescope in Antarctica, could provide a more stringent test soon. “If this really is true — if the early Universe really did feature early dark energy — then we should see a strong signal,” says Colin Hill, a co-author of the ACT team’s paper<sup>1</sup> who is a cosmologist at Columbia University in New York City.

## Mapping the CMB

Both the ACT and the South Pole Telescope are designed to map the cosmic microwave background (CMB), primordial radiation sometimes described as the afterglow of the Big Bang. The CMB is one of the pillars of cosmologists’ understanding of the Universe. By mapping subtle variations in the CMB across the sky, researchers have found compelling evidence for the ‘standard model of cosmology’. This model describes the evolution of a Universe containing three primary ingredients: dark energy; the equally

mysterious dark matter, which is the primary cause of the formation of galaxies; and ordinary matter, which accounts for less than 5% of the Universe's total mass and energy.

Current state-of-the-art CMB maps were provided by the European Space Agency's Planck mission, which was active between 2009 and 2013.

Calculations based on Planck data predict — assuming that the standard model of cosmology is correct — exactly how fast the Universe should be expanding now. But for the past decade or so, increasingly accurate measurements of that expansion, based on observations of supernova explosions and other techniques, have found it to be 5–10% faster<sup>3</sup>.



### [How fast is the Universe expanding? Cosmologists just got more confused](#)

Theorists have suggested a plethora of modifications to the standard model that could explain this difference. Two years ago, cosmologist Marc Kamionkowski at Johns Hopkins University in Baltimore, Maryland, and his collaborators suggested an extra ingredient for the standard model<sup>4</sup>. Their 'early dark energy' — which made more precise an idea that they and other teams had been working on for several years — would be a sort of fluid that permeated the Universe before withering away within a few hundred thousand years of the Big Bang. "It's not a compelling argument, but it's the only model we can get to work," says Kamionkowski.

Early dark energy would not have been strong enough to cause an accelerated expansion, as ‘ordinary’ dark energy is currently doing. But it would have caused the plasma that emerged from the Big Bang to cool down faster than it would have otherwise. This would affect how CMB data should be interpreted — especially when it comes to measurements of the age of the Universe and its rate of expansion that are based on how far sound waves were able to travel in the plasma before it cooled into gas. Planck and similar observatories use features that were left in the sky after this transition to make such calculations.

The two latest studies find that the ACT’s map of the CMB’s polarization fits better with a model including early dark energy than with the standard one. Interpreting the CMB on the basis of the early dark energy model and ACT data would mean that the Universe is now 12.4 billion years old, about 11% younger than the 13.8 billion years calculated using the standard model, says Hill. Correspondingly, the current expansion would be about 5% faster than the standard model predicts — closer to what astronomers calculate today.

## Inconsistencies remain

Hill says that he was previously sceptical about early dark energy, and that his team’s findings surprised him. Vivian Poulin, an astrophysicist at the University of Montpellier in France and a co-author of the second study<sup>2</sup> based on ACT data, says it was reassuring that his team’s analysis agreed with the ACT team’s own. “The lead authors are very, very hard-nosed, conservative people, who really understand the data and the measurements,” Kamionkowski says.

But Galli warns that the ACT data seem to be inconsistent with calculations by the Planck team, which she was part of. And although the ACT’s polarization data might favour early dark energy, it is unclear whether its other major set of data — its map of CMB temperatures — shows such a preference. For these reasons, she adds, it will be crucial to cross-check the results using the South Pole Telescope, an experiment she is part of.

Wendy Freedman, an astronomer at the University of Chicago in Illinois who has contributed to some of the most precise measurements of cosmic expansion, says that the ACT-based results are interesting, if preliminary. “It is important to pursue different models” and compare them with the standard one, she says.

*Nature* **597**, 460-461 (2021)

doi: <https://doi.org/10.1038/d41586-021-02531-5>

## References

1. 1.

Hill, J. C. *et al.* Preprint at <https://arxiv.org/abs/2109.04451> (2021).

2. 2.

Poulin, V., Smith, T. L. & Bartlett, A. Preprint at  
<https://arxiv.org/abs/2109.06229> (2021).

3. 3.

Di Valentino, E. *et al.* *Class. Quantum Grav.* **38**, 153001 (2021).

4. 4.

Poulin, V., Smith, T. L., Karwal, T. & Kamionkowski, M. *Phys. Rev. Lett.* **122**, 221301 (2019).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02531-5>

- NEWS
- 13 September 2021

# Swedish research misconduct agency swamped with cases in first year

The newly formed government organization tackled 46 research-fraud investigations in 2020 — three times as many as expected.

- [Holly Else](#)



The Swedish parliament building in Stockholm. In 2019, Sweden created a government agency to handle research-misconduct investigations. Credit: Alamy

Scientists have inundated Sweden's new national research-misconduct investigation agency with cases, and there is no sign of a let-up in referrals.

Researchers brought 46 cases to the organization — called the National Board for Assessment of Research Misconduct (NPOF) and based in Uppsala — in its first year, according to a report detailing its activities in 2020. This caseload was three times higher than officials were expecting.



### Scandal-weary Swedish government takes over research-fraud investigations

In most countries, universities and research institutions deal with misconduct allegations in-house, which can lead to some cases not being handled fairly or transparently. Sweden followed Denmark — the first country in the world to set up such an agency, in 2017 — in a bid to shake up research-fraud probes.

Experts had warned that the nascent agency could be overwhelmed, and say that the high number of cases could be down to researchers feeling more comfortable about reporting suspicions to an independent agency than to their own institutions, as they did under the previous system.

So far, investigations into 25 of the 46 cases have concluded, with 11 judged to be outside the agency's remit, 10 researchers acquitted and 4 researchers found guilty of misconduct. Last month, the researcher at the centre of the agency's first guilty verdict won her court appeal against the decision.

## Rebuilding trust

Sweden created the agency after [trust in its science was shaken by the case of star surgeon Paolo Macchiarini, formerly at the prestigious Karolinska Institute in Stockholm](#). Macchiarini was eventually found guilty of misconduct relating to trials of an experimental trachea-transplant method, after being cleared by three in-house investigations that were later deemed to be flawed by an independent investigation commissioned by the institute.

Following the scandal, an inquiry led by Margaretha Fahlgren, a literature researcher at Uppsala University, suggested that Sweden establish a government body to handle allegations of serious research fraud — defined as fabrication, falsification or plagiarism — at publicly funded institutions. In 2019, parliament passed a law to define research misconduct and establish the NPOF. The agency began operating in January 2020.

In its first-year report, the NPOF said that 30 of the 46 cases it investigated concerned medicine, health and natural sciences — although it received referrals from all research areas except agricultural science and veterinary medicine. The 46 cases included 56 incidents of alleged misconduct, with 10 relating to fabrication, 18 to forgery, 18 to plagiarism and 10 other matters.



## What universities can learn from one of science's biggest frauds

The organization handed down its first guilty verdict in September 2020, against biomedical scientist Karin Dahlman-Wright, former vice-president at the Karolinska Institute, who took up her post in the wake of the Macchiarini scandal but stepped down in 2019 when misconduct allegations against her surfaced.

The NPOF found that Dahlman-Wright committed research misconduct, with four of seven research papers investigated containing manipulated images. Dahlman-Wright denied the allegations, and appealed her case at the Administrative Court in Uppsala, which upheld her claim in August. Although the articles “contain images that do not correspond to the results that the images are said to show”, the court said in a statement, it ruled that Dahlman-Wright had not been grossly negligent — an essential component of Sweden’s definition of research misconduct.

The NPOF is now preparing to appeal against this decision. The NPOF did not respond to *Nature*’s requests for a comment about the case. Dahlman-Wright declined to comment.

Dahlman-Wright’s appeal was one of two against guilty verdicts by the NPOF. A verdict is still awaited on the other appeal, in a case that involves 13 materials and nanotechnology researchers at Linköping University,

which, the NPOF ruled, fabricated X-ray diffractograms in four research papers. The two other researchers found guilty of misconduct have not appealed and any sanctions will be carried out by their institutions in accordance with the law, a representative of the NPOF told *Nature*.

## Extra staff

Fahlgren, who sits on the NPOF board, says that many cases referred to the agency were the result of personal disputes, particularly between PhD students and their supervisors.

“This is an issue with the work environment — not misconduct — and we hope to communicate with universities about how to deal with this,” says Fahlgren.

For 2021, the NPOF expects to receive a similar number of referrals as in 2020, and it has taken on another staff member to help address the caseload.



### [Check for publication integrity before misconduct](#)

C. K. Gunsalus, a research-integrity specialist at the University of Illinois at Urbana–Champaign, says that the referral figures in Sweden are consistent with what has been seen in the United States for many years. Awareness of responsible research standards and the idea that people are more comfortable

taking concerns to an independent body rather than their own institutions could be behind the unexpectedly high number of referrals, she says.

“It’s past time for the entire research ecosystem to attend to healthy lab cultures at the front end and to provide meaningful, safe and trusted ways to surface issues within institutions — as well as checks and balances for those systems,” Gunsalus adds.

*Nature* **597**, 461 (2021)

doi: <https://doi.org/10.1038/d41586-021-02451-4>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02451-4>

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

- NEWS FEATURE
- 21 September 2021

# How far will global population rise? Researchers can't agree

The United Nations forecasts that nearly 11 billion people will be living on Earth at the end of the century, but other demographic research groups project that population will peak earlier and at a much lower level.

- [David Adam](#) <sup>0</sup>



Illustration by Karol Banach

The 1980s were a puzzling time for would-be parents in Singapore. The government initially told them to “Stop-at-Two” and backed up the policy with a series of measures to deter couples from having three or more children. It increased hospital fees for the delivery of third babies and withdrew maternity pay.

In March 1987, officials performed a demographic U-turn. Under the awkward slogan “Have Three or More (if you can afford it)”, the scales tipped abruptly towards those with larger families, who were now given priority for schools and housing.

Singapore is a dramatic example, but far from unique. Across the world, to secure a stable financial future, governments are desperate to keep national population numbers in a Goldilocks zone: not too many, not too few. And many of these policies are based on computer simulations of how future population numbers will rise and fall.

For decades, the most influential of these projections was produced by a small group of population modellers at the United Nations. But in the past few years, rival groups have developed their own techniques and produced their own results — which vary considerably and have generated bitter disputes in the field.

The [UN says world population will plateau at 10.9 billion](#) by the end of the century. The other groups forecast earlier and smaller peaks, with global population reaching 9.7 billion by 2070 and then declining.

The difference poses a conundrum for governments, companies and others trying to plan for everything from investment in infrastructure and future tax income, to setting goals for international development and greenhouse-gas reductions.



### [The fraction of the global population at risk of floods is growing](#)

No matter which model is used, the most important data are precise numbers of who lives in each country today — and researchers are developing ways to improve these tallies. This is crucial, not only to provide a solid baseline from which to project into the distant future, but also to develop policies for today, such as [allocating COVID vaccinations](#) and providing adequate numbers of school places. And [the pandemic has complicated things](#) by

delaying some censuses and potentially changing predictions for life expectancy and birth rates, at least in the short term.

That adds up to a growing research and policy interest in the planet's human resources.

"Every government is interested in what is going to happen to their population in the next couple of decades, for pragmatic economic reasons and planning needs," says Tomáš Sobotka, a population researcher at the Vienna Institute of Demography.

## Headcount

All estimates of population start with the same question: how many people are alive right now? Attempts to answer that question go back to 4000 bc, when the Babylonians used a census to work out how much food they needed to feed their people. Ancient Egyptian, Roman and Chinese societies all carried out regular counts.

Earlier this year, both [China](#) and [the United States](#) reported results from censuses carried out in 2020. Many more countries were scheduled to release results this year but have been delayed by the ongoing COVID-19 pandemic. "China and the US were very much the exceptions in getting their censuses done," says Patrick Gerland, who leads demographic work at the UN Population Division in New York City, which produces the UN projections.

Both the United States and China reported that they are experiencing record low levels of population growth. Those results made headlines, but they are in line with what demographers expected, says Gerland. That's because both countries track and produce regular, reliable data on births and deaths, which allow population researchers to monitor trends in almost real time.

With census results or other population counts as a baseline, demographers forecast the various ways in which the number of people will change in coming years. Beyond births and deaths, researchers also predict how many people will enter or leave a country over time.

Like all simulations of future events — from climate change to the course of an epidemic — population predictions get less reliable as they are projected over longer time periods. For demographers, projections over the next 20–30 years are usually considered to be very good because most people who will be alive in a few decades' time have already been born. And birth, death and migration rates are fairly easy to extrapolate over that period from recent trends.

These short- to medium-term predictions do remain vulnerable to shocks, however. Demographers are racing to understand the impact of the COVID-19 pandemic, for example. In some of the worst-hit countries, the [large number of deaths in a relatively short period](#) has already had an effect in lowering life expectancy.



The United Nations forecasts that Nigeria's population will more than triple by the end of the century.Credit: Sean Sutton/Panos Pictures

With migration suspended between most countries, the biggest other factor in these post-pandemic calculations of population is the impact on births. Called the fertility rate, the number of children each woman has on average

is a totemic figure in demography. That's because, with death rates and migration usually fairly stable in comparison, large swings in the fertility rate tend to dominate both the actual size of a population and predictions of how that population will shrink or grow in future.

When Singapore, for example, first urged each family to have only two children at most in 1972, the fertility rate in the country was estimated at 3.04 and was forecast to rise sharply. By 1986, just ahead of its policy U-turn and plea for more babies, fertility had plummeted to 1.43. It dipped as far as 1.14 in 2018 and today remains at a worryingly low 1.23.

To maintain a stable population without immigration, a country's fertility rate needs to be at the 'replacement level' of 2.1 births per woman.

Predictions of population in both the near and long-term future typically come down to estimates of how quickly fertility will change. And that means demographers have to make some educated guesses about how people will behave as their circumstances alter. In high-income countries, these behavioural changes are usually driven by economic factors. As opportunities emerge, women prioritize careers, and couples delay having children during a recession.



### [How many people has the coronavirus killed?](#)

In less-wealthy nations, other factors dominate. As more girls are educated, they tend to have fewer children and to have them later. And more people

have access to contraception as health systems and distribution networks improve. In that sense, falling fertility rates reflect economic development.

Demographers expect that the pandemic will cause a short-term dip in fertility, in richer countries at least, because of the associated economic uncertainty. By contrast, poorer countries could see a surge in pandemic births because of the disruption to contraception supplies.

In a preprint<sup>1</sup>, Sobotka and colleagues report on data for 17 countries across Europe, Asia and the United States showing that the number of births did fall — on average by 5.1% in November 2020, 6.5% in December 2020 and 8.9% in January 2021, compared with the same months of the previous year. Spain sustained the sharpest drop in the number of births among the countries analysed, with that metric plummeting by 20% in December 2020 and January 2021 compared with a year earlier.

Some experts predict that births will rebound. “By October it looks like we might be back to normal birth volume,” says Molly Stout, an obstetrician at the University of Michigan Health in Ann Arbor. Over the course of the pandemic, Stout’s team has used electronic health records to model the number of pregnancies in the surrounding community and so plan for an expected number of births. Its published analysis<sup>2</sup> accurately predicted a 14% year-on-year fall in births between November 2020 and March 2021, and suggested a comparable surge in births in the last third of 2021.

## Future faces

How fertility rates and population numbers will change in the longer term, over several decades or more, is more difficult to predict. And this is where the serious controversy starts.

For decades, the UN Population Division had the field largely to itself, and churned out routine updates every couple of years. Its most recent report, published in 2019, forecasts that global population will continue to rise from its current 7.7 billion and [could reach nearly 11 billion by 2100](#). (Its next biennial update was due this year but has been postponed to 2022.)



### IPCC climate report: Earth is warmer than it's been in 125,000 years

In 2014, a group at the International Institute for Applied Systems Analysis (IIASA) in Vienna produced its own forecast. It said that world population is most likely to peak at 9.4 billion around 2070 and will fall to 9 billion by the end of the century. The group's numbers rose slightly in a 2018 report<sup>3</sup> that projected a peak of about 9.8 billion around 2080; a subsequent update has population cresting at a little under 9.7 billion around 2070.

And then, last year, a paper<sup>4</sup> from a team at the Institute for Health Metrics and Evaluation (IHME) at the University of Washington's School of Medicine in Seattle, suggested that global population will peak at around 9.7 billion in 2064, and then decline to about 8.8 billion by 2100.

Some 23 countries could see their current population halved by the end of the century, the study said, including Japan, Thailand, Italy and Spain.

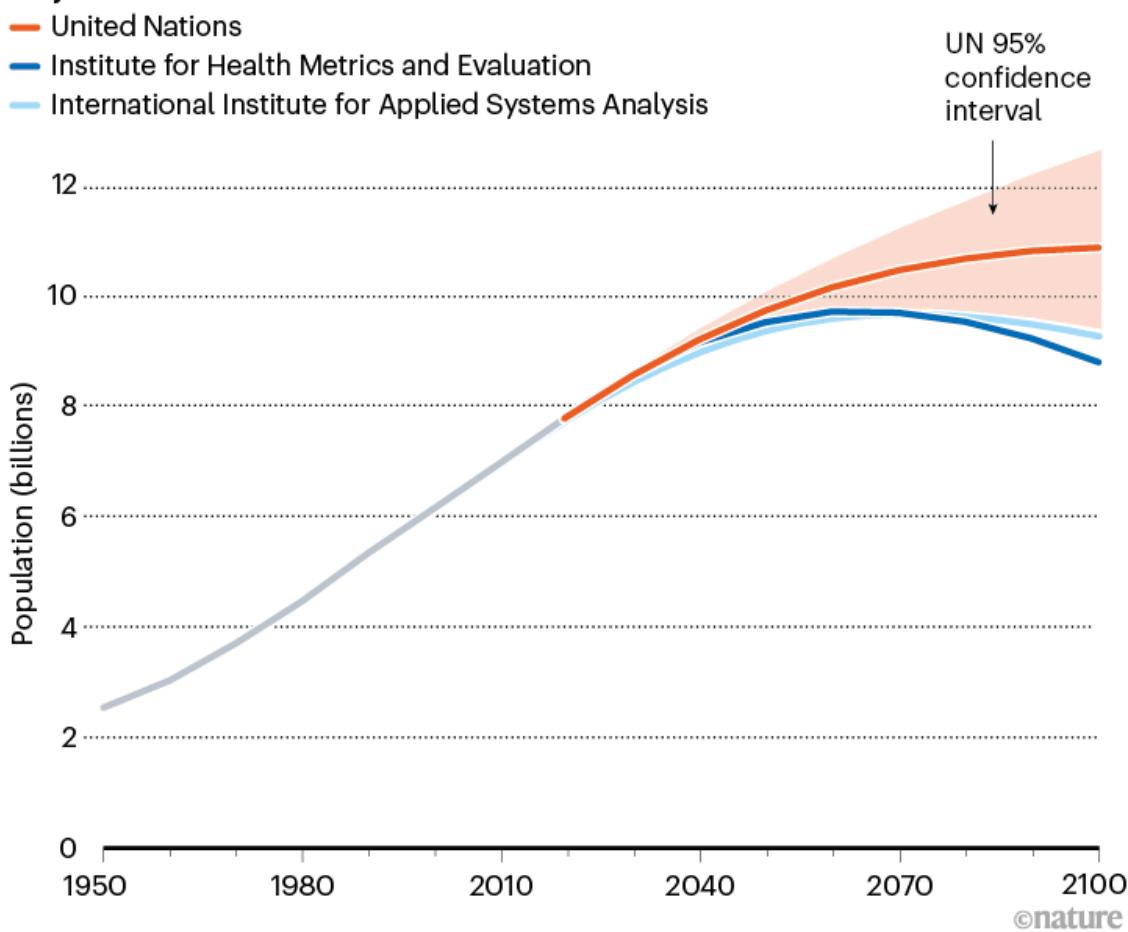
The different outcomes reflect the uncertainty in making projections over such a long time period, says Leontine Alkema, a statistical modeller at the University of Massachusetts Amherst. “It’s kind of an impossible exercise and so we do the best we can and it’s good that different groups use different approaches,” she says.

The huge range between the studies (see ‘Peak people’) comes down to the way each research group sets up its forecast. “All three have used a similar starting point [for population] so we know that’s not the problem,” says Toshiko Kaneda, a demographer at the Population Reference Bureau, an independent research organization in Washington DC. “The issue then is how you think the line will creep up. The assumptions there are where people get it wrong.”

## PEAK PEOPLE

The United Nations projects that global population will reach close to 11 billion by 2100, significantly higher than estimates from two other organizations.

### Projection



Source: UN Population Division/IIASA/IHME

Changes in fertility rates with anticipated economic development are key, and the three models account for this process in different ways. The UN

modellers divide the way that fertility tends to slow, decline and then recover into several phases. Changes in each country's birth rate in recent years are then used to place each nation into one of these phases, and some 100,000 possible pathways for future fertility are modelled. The UN then takes the median of these projections and presents it as the most likely scenario.

Instead of relying on data and past trends to forecast future falls in fertility due to development, the IIASA group turned to expert judgement. They asked some 200 researchers, including economists, demographers and sociologists, to predict fertility rates for individual countries in 2030 and 2050, on the basis of what they expected to happen to several social, health and economic factors. Some of these estimates varied considerably. Forecasts of fertility rates in India ranged from 1.5 to 2.5 for 2030 and from 1.1 to 2.5 for 2050.

IIASA's fertility-rate forecasts are noticeably lower than the UN's. For example, the researchers estimate that fertility for every country in sub-Saharan Africa will be below the replacement level of 2.1 by the end of the century. The UN forecasts that only one-third of countries in the region will fall below this level.

The IHME team did things quite differently. Instead of basing its model on fertility rates, and how they would change, the group used a variable called completed cohort fertility at 50 years (CC50). This counts the number of children each woman has had by the time she reaches 50. It is subtly different from overall fertility rate because it is less sensitive to the age at which women have their children, and it does not show the same rebound effect when fertility drops to low levels.

And instead of assuming a figure for this CC50 at specific time points in the future, the IHME model used real-world data to work out the relationship between CC50 and its two main drivers: educational attainment and unmet contraceptive need. This meant they could plug national data on education and contraception — and how they expected them to change — into the model instead of simple estimates of future fertility.



## [COVID vaccines to reach poorest countries in 2023 — despite recent pledges](#)

Christopher Murray, who leads the IHME team, says this approach makes the IHME forecast more reliable and more valuable because it can test the impact of changes and assumptions. The model could forecast what happens to population when policies encourage more girls to spend longer in school, or when health infrastructure improves to secure more reliable access to contraception. “In the policy realm it’s much more useful to have models with causal connections,” Murray says.

Plenty of demographers disagree. “There are a few issues with those [IHME] projections that are a little bit problematic,” says Stuart Gietel-Basten, a demographer at the Hong Kong University of Science and Technology. Together with Sobotka in Vienna, Gietel-Basten published a preprint<sup>5</sup> of a technical critique of the IHME study that highlights what they claim are “internal inconsistencies, discrepancies and illogical and implausible trends”.

For example, the duo points to Iraq, which the IHME forecasts will boast the world’s fourth-highest female life expectancy by 2100, as well as welcoming huge numbers of immigrants in the coming decades. The critics say this is highly unlikely. Sobotka and Gietel-Basten have organized a critical letter signed by 170 demographers and sent it to *The Lancet*, which published the IHME paper. The letter has not yet been published.

“The big concern I have is that a projection can shape the future,” says Gietel-Basten. “If you’re going to say we are going to have very, very low fertility, rapid population ageing and stagnation, well that’s not what governments want.” As seen in Singapore, politicians can react with policies designed to prevent or, more commonly, produce more babies. “They can react by restricting access to family planning, restricting access to abortion and restricting access to vasectomies.”

Kaneda says that the IHME group that produced the paper has little background in demography. Instead, it based its population forecast on methods it developed to calculate a regular set of health statistics called the Global Burden of Disease. “I think it’s a great effort, just that they should go back and revise some of this stuff,” says Kaneda.

Murray rejects the criticism, saying that the UN model itself carries its own “strange set of assumptions” and that the demography community is reluctant to accept ideas from outside the field. “Let’s look at how forecasts go in the next five or ten years in places with low fertility,” he says. “Are we going to see in China and Korea and Singapore, Greece and Spain that fertility shoots up as the UN says, or not? I think we won’t.”

The UN’s past forecasts have a decent track record. In 1968, for example, the UN projected that the global population in 1990 would be 5.44 billion — within 2% of the best estimate of the real figure of 5.34 billion. In 2010, the estimated global population was 7 billion, compared with projections in previous UN reports that ranged from 6.8 billion to 7.2 billion.

The organization is also using new and better sources of data about populations in specific countries to upgrade its historical records, Gerland says. This will make the modelling more accurate, he adds, and should allow for more regular updates — although the current update is taking longer than expected and has delayed the latest global population report.

## Right here, right now

Some demographers stay on the sidelines. “I steered well clear of getting involved in any of that because it got quite nasty and it’s very difficult to

really say what's the better approach," says Tom Wilson, a demographer at the University of Melbourne, Australia. "The one thing unfortunately about population projections is they will always turn out to be wrong."



### COVID boosters for wealthy nations spark outrage

That's why some in the field prefer to leave the future alone and focus instead on improving the accuracy of data used immediately to set policy: counting people alive right now. In some places, especially those facing instability and civil strife, that's more difficult than it sounds. "In Afghanistan, the last census was in 1979. In the DRC it was 1984," says Andy Tatem, a population researcher at the University of Southampton, UK. In those cases, governments tend to assume a linear annual increase to estimate current numbers. But that could be wildly inaccurate. A 2017 analysis<sup>6</sup> by researchers at the University of Antwerp in Belgium found that national population estimates used by the government of the Democratic Republic of the Congo ranged from 77 million to 102 million.

To produce better data, researchers are testing ways to count people without actually counting them.

One technique is to monitor mobile-phone traffic. By tracing calls to the phone towers that send and receive them, researchers can use call density around the towers to estimate the local population. In one high-profile

application of this technique, researchers from Sweden and South Korea tracked the displacement of people after a devastating earthquake struck Haiti in 2010. The research showed that the population of the capital, Port-au-Prince, shrunk by almost one-quarter within three weeks of the quake<sup>7</sup>.

Tatem's team has applied a similar technique to Namibia in a study of malaria prevalence in different parts of the country. The results suggested that Namibia was closer to eliminating the disease than policymakers realized at the time.

Researchers are also working to count people on the basis of the size and shape of the buildings they live in. Using satellite photos and image-recognition software, they can map settlements and individual houses, and then build up a picture of the number of residents. “We’ve done this to fill in gaps in the Colombia census and the Burkina Faso census, and to produce new estimates for the DRC and Zambia and quite a few other countries,” Tatem says. “It’s an approach that is starting to take off.”

Even so, old-fashioned population counts still have their place. “The census collects so much more than just population numbers,” Tatem adds. “These methods should be seen as a complement to the census rather than something to replace it.”

*Nature* **597**, 462-465 (2021)

doi: <https://doi.org/10.1038/d41586-021-02522-6>

## References

1. 1.

Sobotka, T. *et al.* Preprint at SocArXiv  
<https://osf.io/preprints/socarxiv/mvy62> (2021).

2. 2.

Stout, M. J. *et al.* *JAMA Netw. Open* **4**, e2111621 (2021).

3. 3.

Lutz, W., Goujon, A., Samir, K. C., Stonawski, M. & Stilianakis, N. (eds) *Demographic and Human Capital Scenarios for the 21st Century* (European Union, 2018).

4. 4.

Vollset, S. E. *et al. Lancet* **396**, 1285–1306 (2020).

5. 5.

Gietel-Basten, S. & Sobotka, T. Preprint at SocArXiv  
<https://doi.org/10.31235/osf.io/5syef> (2020).

6. 6.

Marivoet, W. & De Herdt, T. *From Figures to Facts: Making Sense of Socioeconomic Surveys in the Democratic Republic of the Congo (DRC)* (Institute of Development Policy and Management, 2017).

7. 7.

Lu, X., Bengtsson, L. & Holme, P. *Proc. Natl Acad. Sci. USA* **109**, 11576–11581 (2012).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02522-6>

# How dangerous is Africa's explosive Lake Kivu?

An unusual lake in central Africa could one day release a vast cloud of greenhouse gases that suffocates millions of people. But it's not clear whether the threat is getting worse.

**By Nicola Jones**  
**23 September 2021**



Credit: Guerchom Ndebo for *Nature*

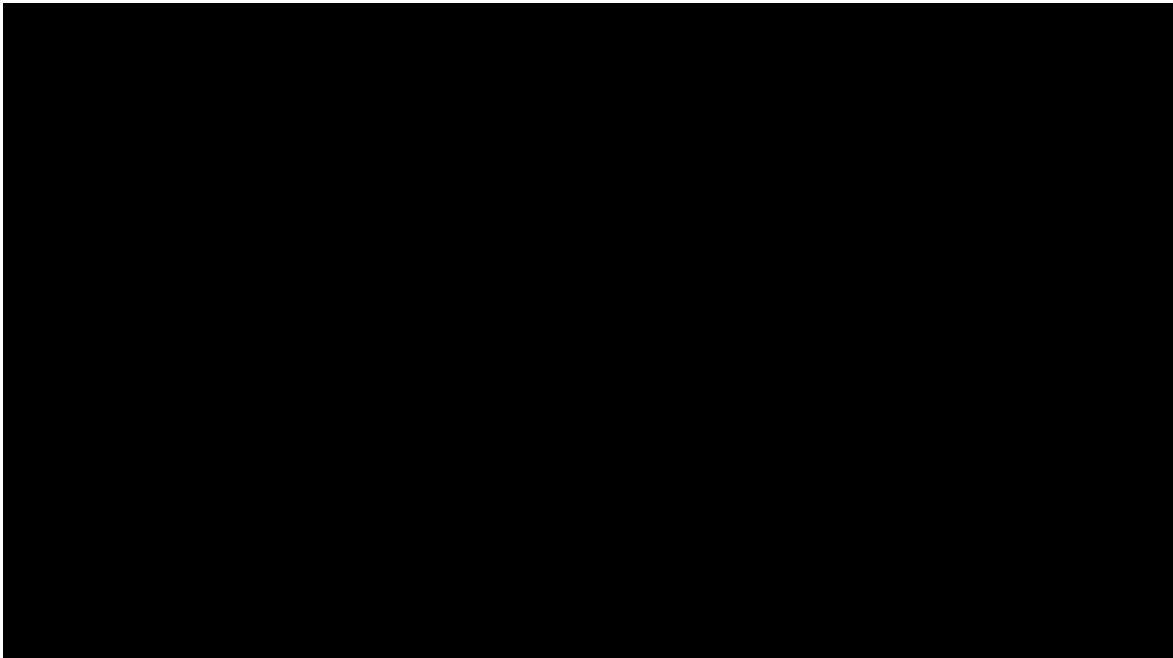
Credit: Guerchom Ndebo for *Nature*

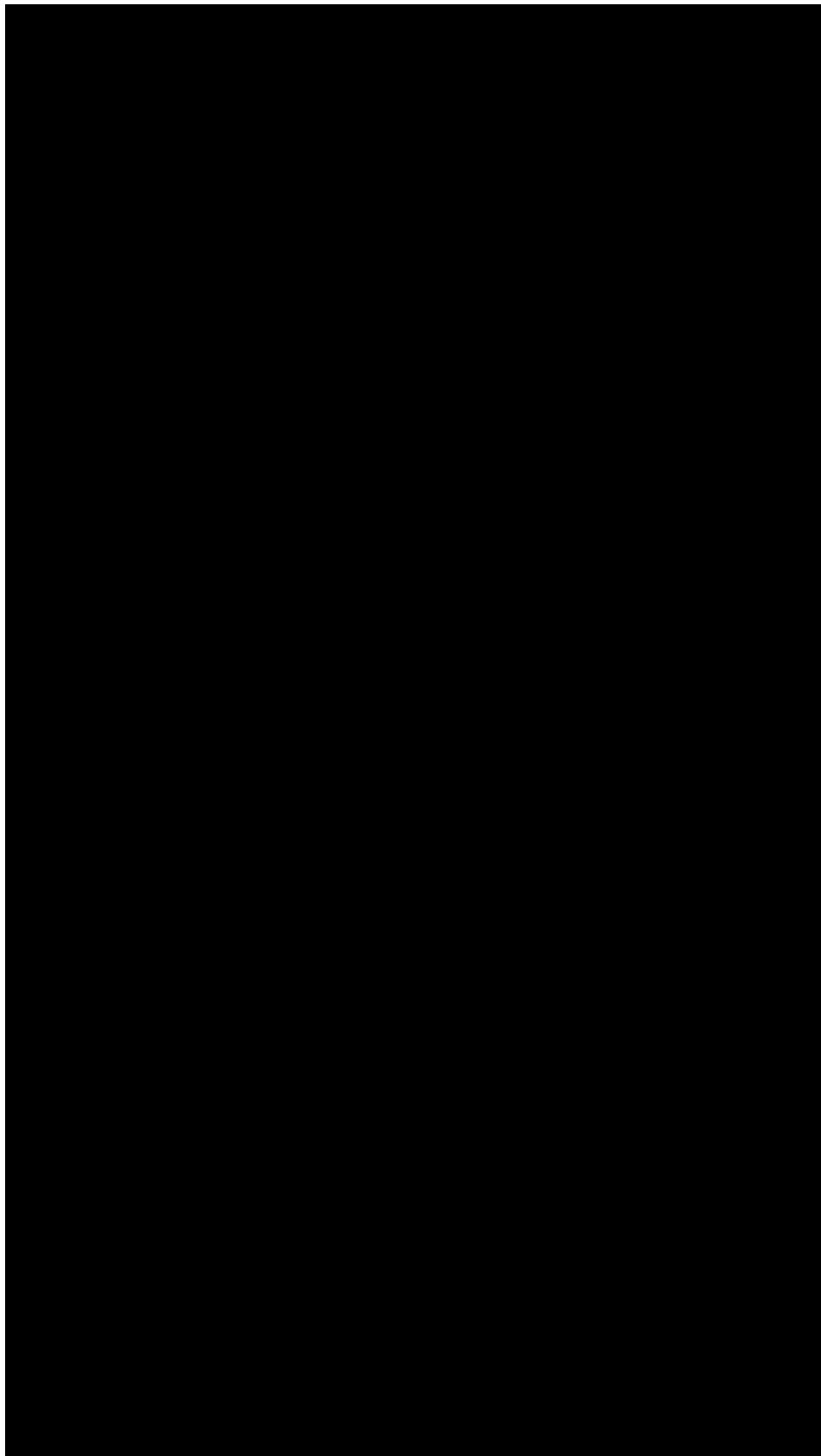
On 22 May, one of Africa's most active volcanoes, Mount Nyiragongo, started spewing lava towards the crowded city of Goma in the Democratic Republic of the Congo (DRC). The eruption destroyed several villages, killed dozens of people and forced an estimated 450,000 people to flee their homes.

The volcano has since calmed and the immediate humanitarian crisis has eased. But government officials and scientists have another worry on their minds: something potentially even more dangerous than Mount Nyiragongo.

Goma sits on the shore of Lake Kivu, a geological anomaly that holds 300 cubic kilometres of dissolved carbon dioxide and 60 cubic kilometres of methane, laced with toxic hydrogen sulfide. The picturesque lake, nestled between the DRC and Rwanda, has the potential to explosively release these gases in a rare phenomenon known as a limnic eruption. That could send a huge pulse of heat-trapping gases into the atmosphere: the lake holds the equivalent of 2.6 gigatonnes of CO<sub>2</sub>, which is equal to about 5% of global

annual greenhouse-gas emissions. Even worse, such a disaster could fill the surrounding valley with suffocating and toxic gas, potentially killing millions of people. “It could create one of the worst, if not the worst, natural humanitarian disasters in history,” says Philip Morkel, an engineer and founder of Hydragas Energy, based in North Vancouver, Canada, who is attempting to get funding for a project to remove and utilize gas from the lake.







Mount Nyiragongo erupted in May and displaced nearly half a million people in Goma. Credit: Alex Miles/AFP via Getty

---

The 2021 volcanic eruption didn't trigger a mass release of gases from the lake, and on 1 June, the Rwanda Environment Management Authority (REMA) said there was no imminent risk. But, the authorities do think that lava flowed through underground fractures beneath the city of Goma and Lake Kivu itself. A day after the eruption, a tremor seems to have triggered part of a sandbar by the lake to collapse, which might have caused a small release of gases in that spot: some people reported that waters offshore from a prominent hotel looked like they were boiling.

For now, the lake is stable. Although it contains a lot of gas, the concentration would have to double in the region with the most gas for it to reach saturation point. But a strong earthquake or volcanic eruption could potentially trigger a gas release by disrupting the lake's layered structure or increasing the gas concentrations. And some researchers worry that a disaster might be brought on by human activity, too.

Methane is already being pumped from the lake's depths and burnt to create much-needed electricity, which most people agree is both a sensible use of

local natural resources and a way to make the lake safer by removing some of its gas. The stakes are high: researchers have estimated that the methane in Lake Kivu could be worth up to US\$42 billion over 50 years.

But researchers disagree about which method of gas extraction is best, and whether such efforts might eventually disturb the lake in ways that elevate the dangers rather than subduing them. The debate rages even while efforts to harvest methane are expanding — plans are in place to bump up electricity generation more than fivefold in the coming years or decades.

“A lot of scientists don’t agree,” says biochemist Eric Ruhanamirindi Mudakikwa, head of the Rwandan government’s Lake Kivu Monitoring Program (LKMP). “What we are doing on the lake is really new,” he says. “We don’t know how it can behave.”

Residents of Goma in the Democratic Republic of the Congo are flanked by Nyiragongo volcano and Lake Kivu, both of which pose threats.

A gas eruption from Lake Kivu could threaten millions of people living in the region.

Credit: Guerchom Ndebo for *Nature*

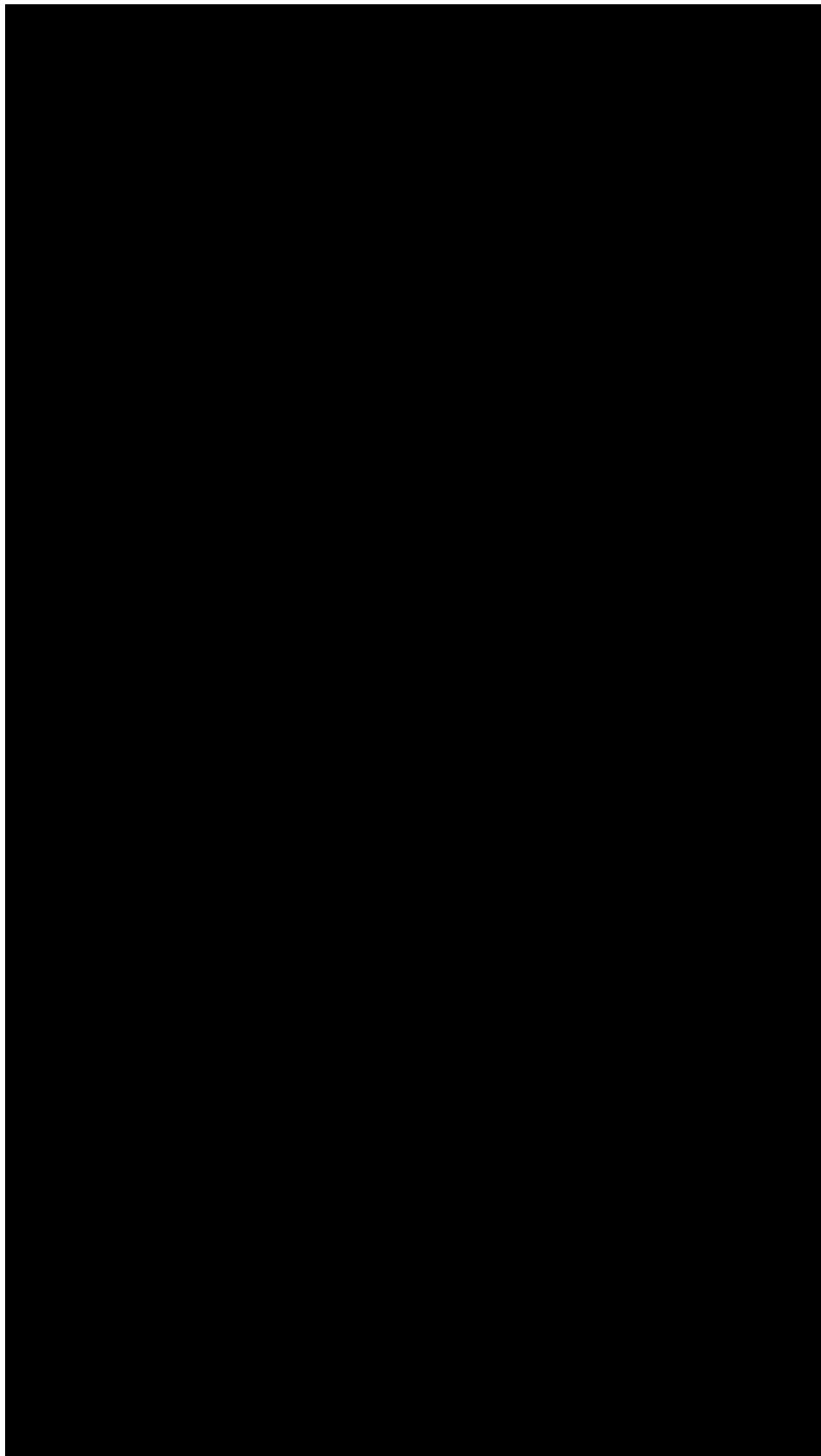
---

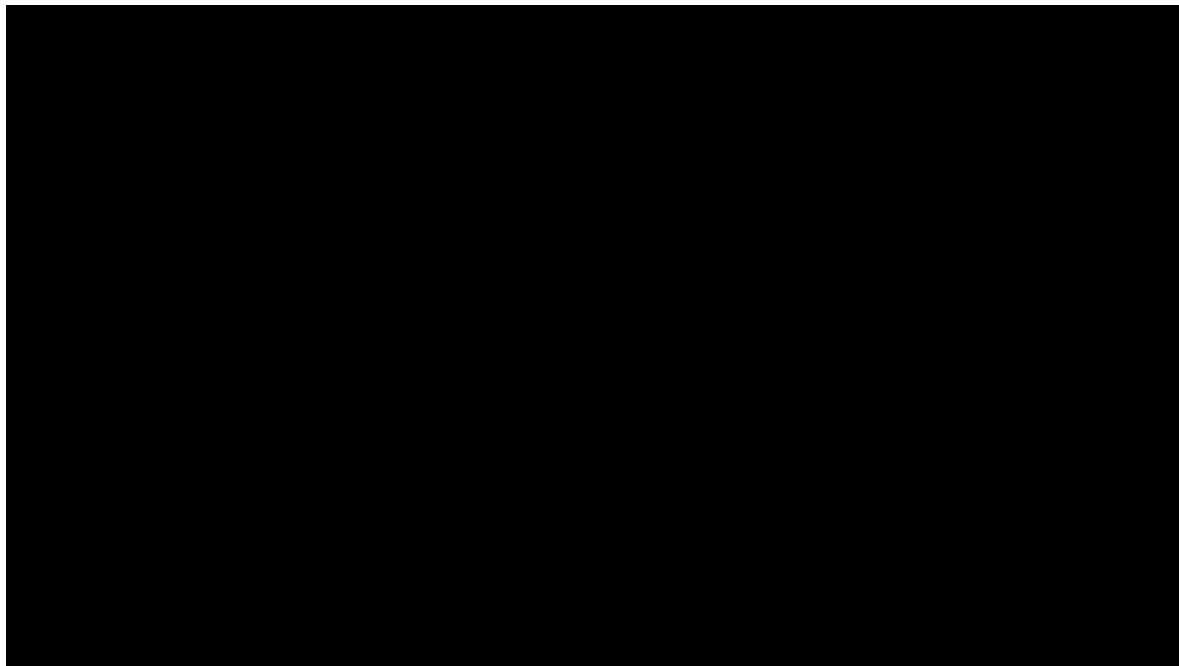
## Under Pressure

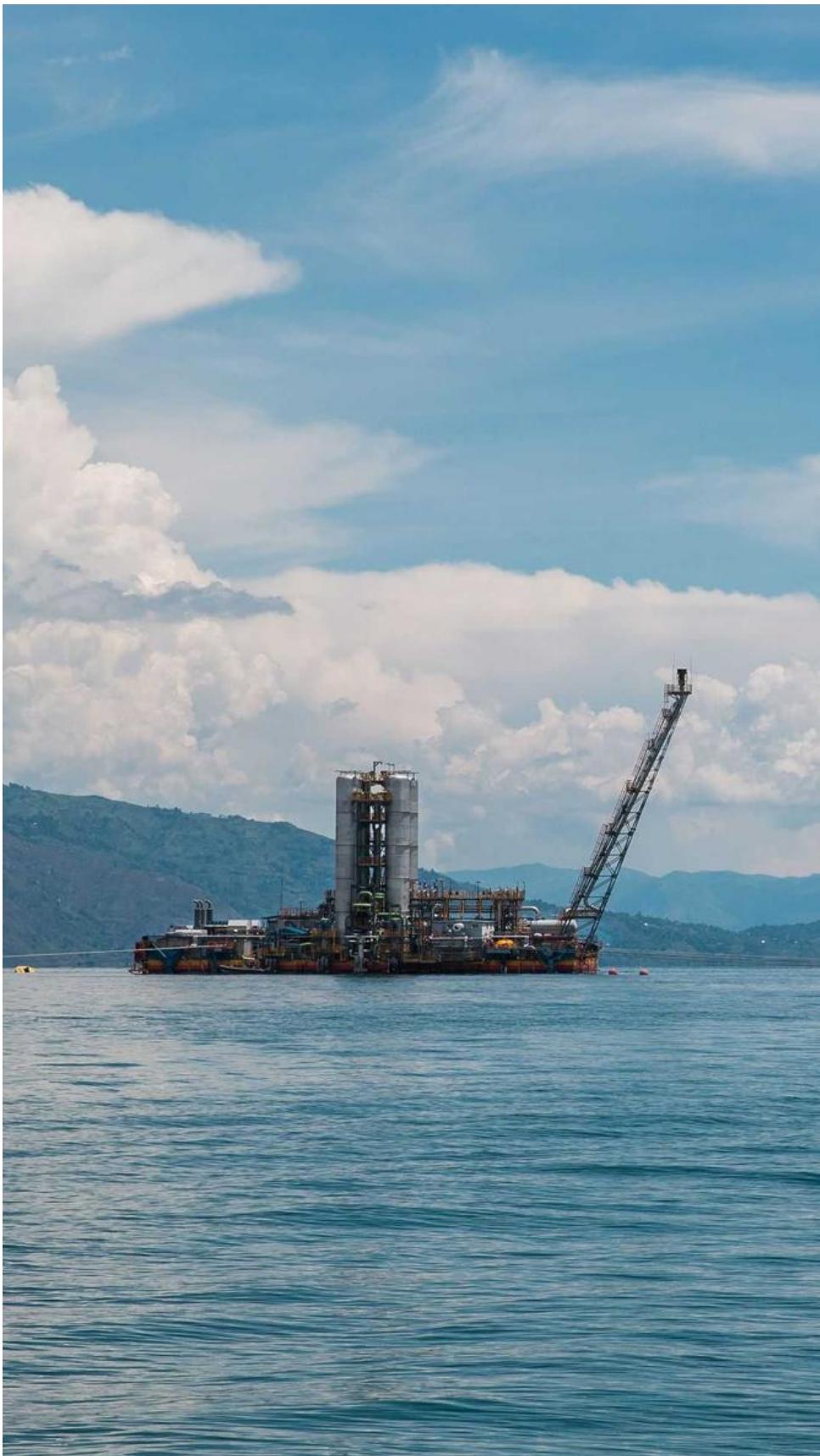
Lake Kivu is the largest of only a handful of lakes in the world thought to be capable of limnic eruptions. Two, much smaller, such lakes lie thousands of kilometres west, in Cameroon; and another, Lake Albano, is in Italy.

These lakes all sit above tectonically active regions, where volcanic gases such as CO<sub>2</sub> seep upwards from deep within Earth. The lakes are deep, and their waters do not mix top to bottom with seasonal temperature swings. Instead, the dissolved gas accumulates in denser bottom layers, capped by a ‘cork’ of pressure from the waters above (see ‘Deep gas’). If the gases accumulate to such an extent that they form bubbles, these lakes can literally explode like a champagne bottle. An external event can also ‘pop the cork’

— a drought could lower lake levels and reduce pressure on the gassy waters below; a landslide, earthquake or lava erupting into the bottom of the lake could shift the water layers or add enough heat to cause gas to bubble out.







The KivuWatt project generates 26 MW of electrical power by extracting methane gas from the lake. Credit: Rachel Couch

The KivuWatt project generates 26 MW of electrical power by extracting methane gas from the lake. Credit: Rachel Couch



The KivuWatt project generates 26 MW of electrical power by extracting methane gas from the lake. Credit: Rachel Couch

The violent potential of these lakes became clear in August 1986, when Lake Nyos in Cameroon erupted with a blast that some locals mistook for the testing of a nuclear weapon. As much as 1 cubic kilometre of heavier-than-air CO<sub>2</sub> flooded low-lying regions, suffocating more than 1,700 people and 3,500 livestock.

After the blast, a project was initiated to ensure this wouldn't happen at Lake Nyos again: in 2001, physicist and engineer Michel Halbwachs, then at the University of Savoie in Chambéry, France, and his team inserted a pipe into the lake from a floating dock and siphoned up deep, gassy waters. This created a self-powered fountain, allowing gas to vent in a tiny, controlled version of a limnic eruption. The [team added another two pipes in 2011](#); by

2019, Halbwachs and his colleagues considered Lake Nyos “quite totally emptied of hazardous amounts of dissolved carbon dioxide”<sup>1</sup>.

Halbwachs then tackled Nyos’s little sibling, Lake Monoun, which had experienced a much smaller eruption in 1984. After the venting pipes were installed, the lake was considered degassed by 2009.

Halbwach’s company, Limnological Engineering, has just secured a \$5-million contract to degas CO<sub>2</sub> from the Gulf of Kabuno, a small offshoot at the north end of Lake Kivu, which has high concentrations of CO<sub>2</sub> at shallow depths. The company has had a pilot project under way since 2017.

But the vastly larger Lake Kivu presents a different problem. Lake Kivu is geologically older than Lake Nyos, and the soil surrounding it is richer in organic matter. Unlike at Lake Nyos, this has led to substantial amounts of methane in Lake Kivu, says biogeochemist George Kling at the University of Michigan in Ann Arbor, who studies limnic eruptions. Microorganisms digesting organic matter produce methane, and volcanically produced methane or hydrogen could be seeping directly into the lake from the rocks below. Methane is much less soluble than CO<sub>2</sub>, and so is much closer to bubbling out. “It’s the methane that’s the problem. It’s not like Lake Nyos,” says Alfred Johny Wüest, a lake physicist at the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) in Kastanienbaum.

Although the lake contains a lot of CO<sub>2</sub>, it could safely hold much, much more if the methane wasn’t adding to the gas pressure. Extract the methane for fuel use, and the CO<sub>2</sub> becomes a non-issue, scientists say.

Credit: Nik Spencer/*Nature*

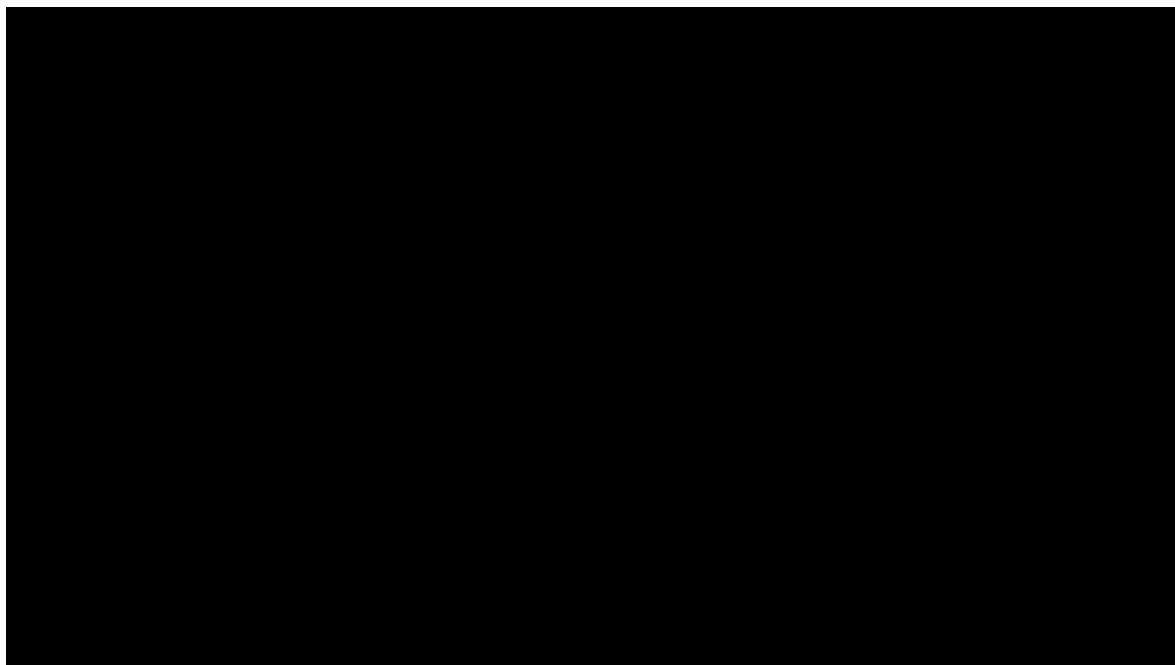
---

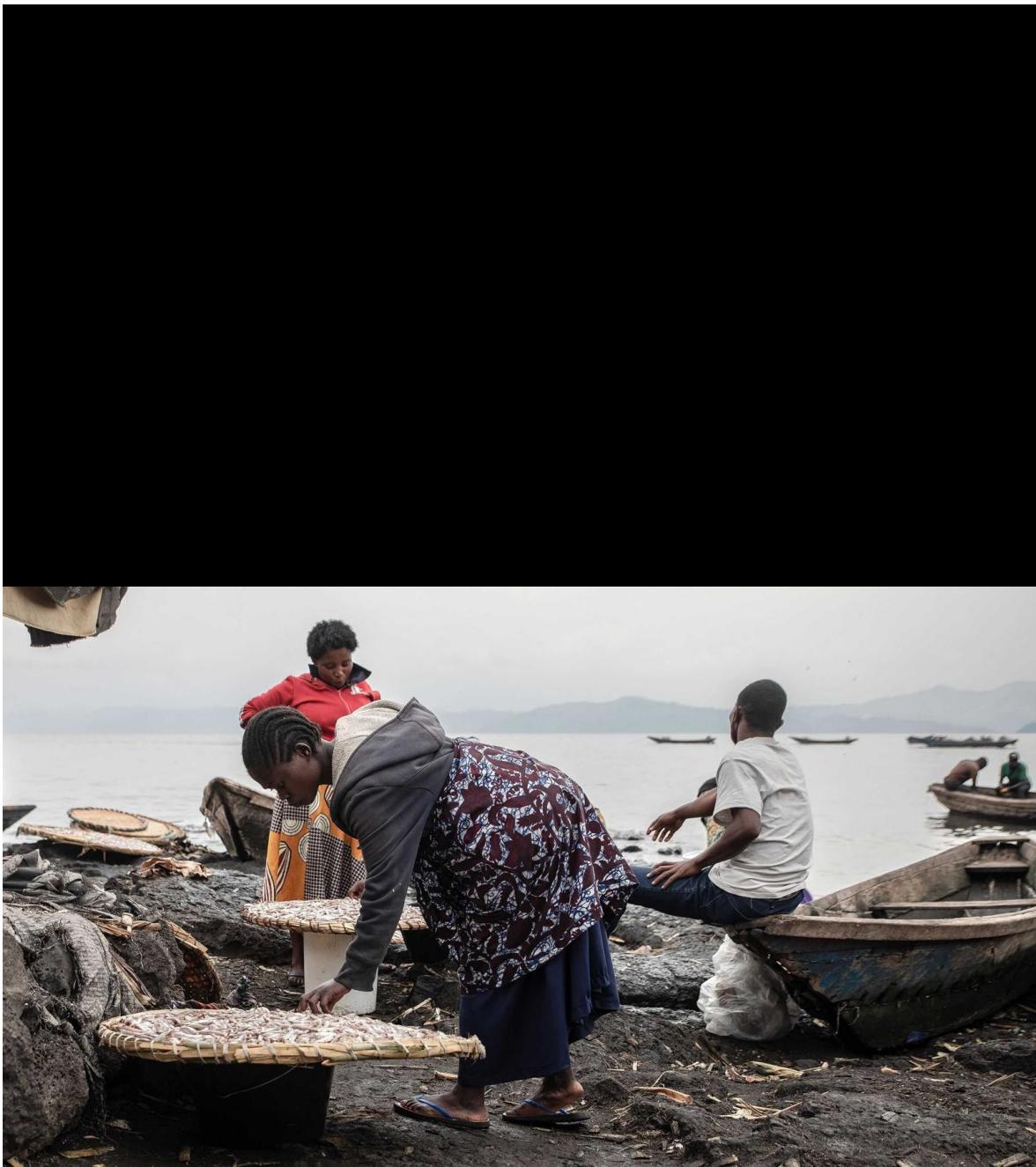
## Gas mysteries

Despite the threat that Kivu potentially poses, there is considerable disagreement on basics, such as the source of the gases, whether amounts are increasing, and even whether Lake Kivu has erupted before. Robert Hecky, a retired lake ecologist at the University of Minnesota Duluth, who has studied

Lake Kivu, says that although there are 9 brown layers in the sediments, showing mixing events in the past 2,000 years, he has found no evidence of any events in the past 12,000 years violent enough to be called a limnic eruption<sup>2</sup>. Others interpret the evidence as signifying at least one eruption 4,000 years ago<sup>3</sup>.

Some facts are clear. The lake's surface waters are fresh and filled with fish. Around 260 metres down, there's a dramatic shift to waters that are much warmer and saltier, thanks to hydrothermal springs. These are the deep 'resource waters' flush with dissolved gas.





A fish seller checks on a haul of fingerlings from Lake Kivu before taking them to Goma. Credit: Guerchom Ndebo for *Nature*

A fish seller checks on a haul of fingerlings from Lake Kivu before taking them to Goma. Credit: Guerchom Ndebo for *Nature*

In 2005, a paper<sup>4</sup> by EAWAG environmental scientist Martin Schmid and his colleagues, including Halbwachs, compared gas levels in that deep layer

with measurements taken in 1975, and suggested that methane concentrations had increased by 15%. If that trend were to continue, the deeper layers would reach saturation by 2090, triggering an eruption. In 2020, however, data in another paper<sup>5</sup> — with Schmid as co-author — suggested the gas levels had not increased after all.

This reassured many researchers, but the findings remain controversial. For one thing, the gas-measurement technique had changed from one data set to the next. “From a methodological standpoint, they are mostly comparing apples to oranges,” says Kling. And the errors on such measures can be large, he says. From Kling’s perspective, the 2020 paper doesn’t prove there has been no change over time, but rather that a change can’t be detected one way or another. “That is a very different thing,” he says.

Whether gas levels have gone up or not, their future is also uncertain — and concentrations could still rise dramatically without warning. “The underground plumbing of the volcanic system of the rift that surrounds Lake Kivu is very poorly understood,” says Kling. “It is quite possible that changes in gas inputs could increase dramatically, due to a rise in subterranean volcanic or geologic activity.”



A team from the Goma Volcano Observatory surveys the crater of Mount Nyiragongo three weeks after its May eruption. Credit: Alexis Huguet/AFP via Getty

A team from the Goma Volcano Observatory surveys the crater of Mount Nyiragongo three weeks after its May eruption. Credit: Alexis Huguet/AFP via Getty

Those same volcanic eruptions and earthquakes could also theoretically trigger an eruption. “You have a gas-rich lake sitting next to a volcano; you have a potential for many triggers,” says Hecky. The question is how big they would have to be. “The lake is exceptionally stable; it would take an enormous amount of energy to overturn it,” he says. Dario Tedesco, a volcanologist at the Luigi Vanvitelli University of Campania, Italy, who works in Rwanda, says his data show that the 2021 volcanic eruption didn’t release gases from fissures around Goma or the lake: magma was either not present underground, he says, or its flows were so small or deep that they had no impact.

Yet most of the dozen or so scientists contacted by *Nature* remain concerned about the lake’s methane levels, given the area’s geological activity. Efficiently extracting 90% of the methane over some 50 years, argues Morkel, could reduce the likelihood of a limnic eruption by 90% in the first 10 years. “In the best case, it will never happen,” he says.





The KivuWatt project in Lake Kivu. Credit: Rachel Couch

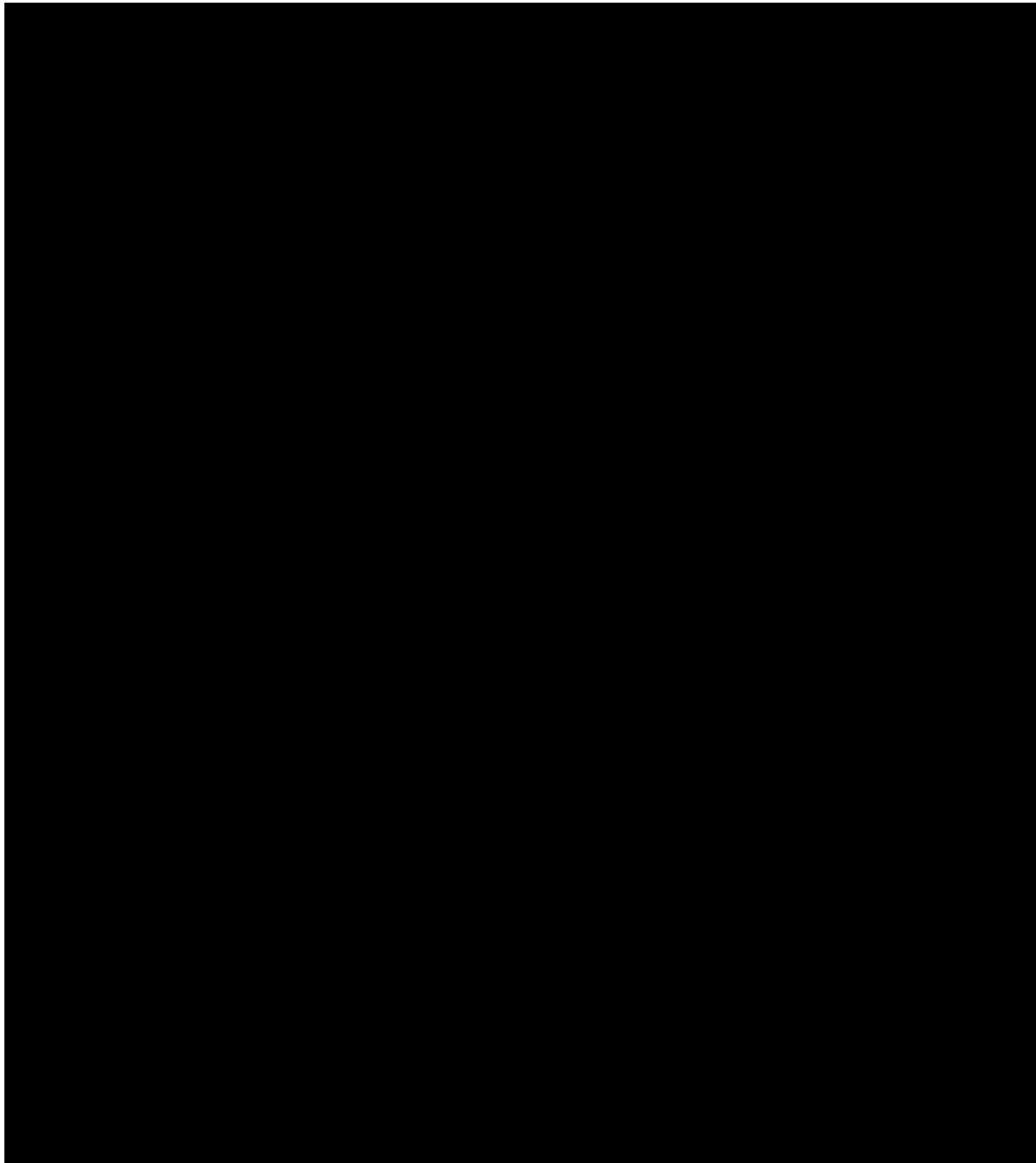
The KivuWatt project in Lake Kivu. Credit: Rachel Couch

## Tapping the methane

People have been pumping methane from Lake Kivu on a small scale for decades to make use of it for energy. But efforts ramped up seriously when KivuWatt, run by London-based ContourGlobal, began operation in 2016. The \$200-million project is currently providing 26 MW of electrical power, and it has a contract to increase that to 100 MW. This will add considerably to Rwanda's baseline installed grid capacity of about 200 MW.

For now, KivuWatt's withdrawals are minor in terms of the lake's stock: at the current rate of extraction, the company will remove less than 5% of the methane in the lake in 25 years. "For sure, this speed cannot be considered sufficient to really decrease the risk of limnic eruption," says Francois Sarchambeau, a limnologist at KivuWatt. "So, we need to expand to more capacity." But expansion plans are on hold until electricity demand catches up with supply, the company says. KivuWatt is also considering options for removing CO<sub>2</sub> from the lake and selling it as a commercial product.

Meanwhile, Rwandan company Shema Power Lake Kivu has bought a tiny pilot plant, KP-1, that started pulling methane from the lake in 2006. The firm is currently constructing a facility planned to deliver 56 MW. The company's website says it expects to have construction finished in early 2022, but Shema Power's project director Tony de la Motte declined to answer *Nature*'s questions about the plant's schedule or details of its operation.





The KivuWatt project pumps methane gas to an onshore power plant where it is burned to produce electricity. Credit: Rachel Couch

The KivuWatt project pumps methane gas to an onshore power plant where it is burned to produce electricity. Credit: Rachel Couch

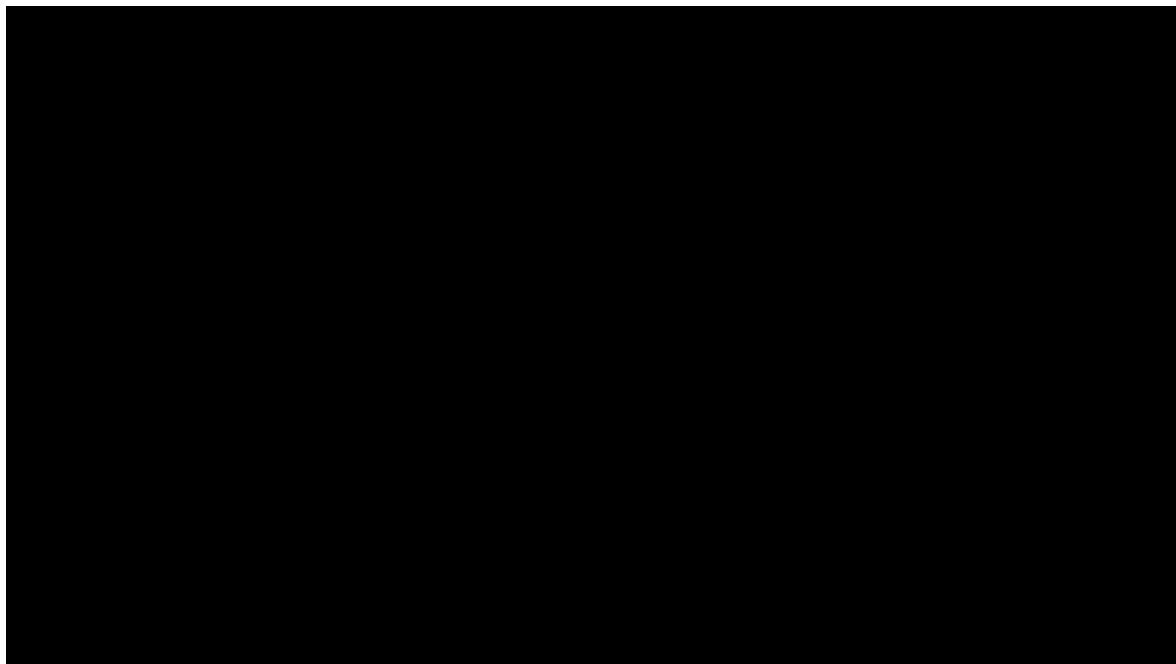
The general principle of all such projects is to pull up deep water so the methane bubbles out and can be purified and pumped to a power plant. The degassed water is then returned to the lake. Questions surround how best to do this; plans vary, depending on the company and the proposal.

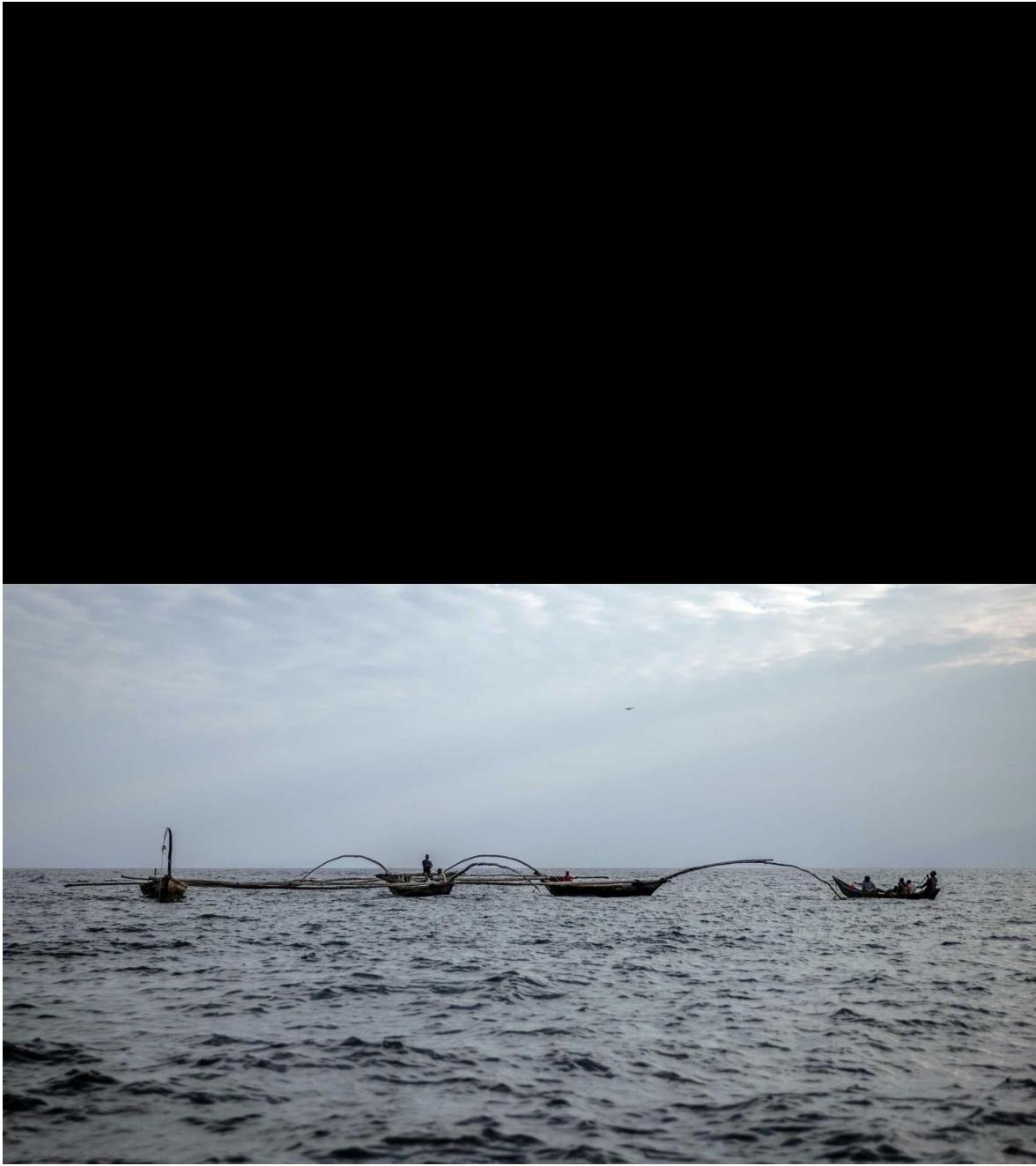
The degassed water still contains high levels of nutrients and toxic hydrogen sulfide, so returning it too near the surface could kill fish and lead to harmful algal blooms, say some researchers. It is also salty and laden with CO<sub>2</sub>, making it relatively dense. So, if released into the lake at too shallow a depth, the degassed water would sink, potentially disturbing the main density gradient, 260 metres deep, that keeps the gassy waters of the resource zone trapped below. “It wouldn’t necessarily blow up, but it would be more prone to blow up,” says Morkel.

Pushing the main gradient upwards could also be problematic, because it would reduce the pressure on the gassy waters. And diluting the resource

layer with degassed water might lower gas concentrations enough that commercial extraction would no longer be possible. If that happened, it would leave a lot of dangerous gas in the lake, with no good way to remove it other than venting it to the surface — an approach that could both release potent greenhouse gases and contaminate surface waters.

In 2009, an international group of researchers, including Morkel, Wüest and Schmid, published '[management prescriptions](#)' (MPs) outlining best practices for extracting the lake's methane. The majority of the experts favoured a strategy called the density zone preservation method, which involves controlling the density of degassed waters by managing the amount of CO<sub>2</sub> they contain, so they can be carefully returned to the lake without causing mixing. This is technically difficult to do, but would largely maintain the current structure of the lake.





A group of fishers take their dugout canoes into Lake Kivu in September.  
Credit: Guerchom Ndebo for *Nature*

A group of fishers take their dugout canoes into Lake Kivu in September.  
Credit: Guerchom Ndebo for *Nature*

KivuWatt opted for an alternative strategy, in which degassed waters are released just above the main gradient. This is simpler to accomplish and

should avoid diluting the resource layer, but is expected to alter the structure of the lake.

Sarchambeau says KivuWatt monitors the surface waters daily, and does weekly profiling to get a robust data set regarding the lake's stability. He says that after five years of operation, the firm did start to see, as expected, a weakening of lake stability — but not by much. "If we pursue the gas extraction as we do, during 50 years we will reduce the lake stability by 1%," he says. This is well below the MPs' guideline, which is that the stability — expressed in terms of the energy needed to completely mix the lake — must not be reduced by more than 25%.

Some argue, however, that KivuWatt's approach is problematic. "That is the way to disaster," says Finn Hirslund, an engineer with consultancy firm COWI, based in Lyngby, Denmark, who was part of the group that wrote the MPs and who has published peer-reviewed papers about Lake Kivu.

Hirslund argues that the project will "destroy the main gradient", and worries that continuing and scaled-up extraction from the lake using similar methodologies might have long-term consequences that only become apparent after decades<sup>6</sup>.

Morkel, too, is critical of KivuWatt's approach. He argues that the company's degassed water has too much CO<sub>2</sub> and is too dense, which he thinks will punch a hole through the main gradient. Morkel advocates taking water and returning it to different depths from those chosen by KivuWatt. He thinks that would better preserve the lake's layering while extracting gas for energy. He continues to try to raise funding for his approach.

Others are not concerned, however. "In terms of safety, I'm absolutely confident," says Wüest, who also serves on KivuWatt's independent expert advisory group. "I have a really positive view on the whole thing," says Bertram Boehrer, a physicist at the Helmholtz Centre for Environmental Research in Magdeburg, Germany, who has worked on the lake. "If something goes in an unexpected way, there's enough time to act."

## Future Forecasts

Perhaps the only way to resolve debate about how these operations might affect the lake is to track whether and how the density layers are changing. The LKMP surveys the depths and inspects the gas-extraction companies, and Mudakikwa says its weekly profiling shows the lake remains stable for now. “The main gradient is not changing,” he says. “If there is a lake instability, we will be the first ones to be concerned.”

KivuWatt says it is required to and does comply with the guidelines set out by the LKMP, and that the company’s independent expert advisory group (including Hecky and Wüest) has access to its data and reviews its annual report to the government of Rwanda.

“We are very open to science,” Sarchambeau says, although some information — such as the design of KivuWatt’s bespoke gas concentration sensors — remains proprietary. “Everyone wants the data from KivuWatt,” says Priysham Nundah, director of KivuWatt. “I cannot give a competitor things,” he says, “But what we are supposed to give [to the LKMP] contractually and based on our obligation, we are doing.”

Some researchers contacted by *Nature* complained that they have had trouble getting access to such data. “In our [MP] guidelines we stated very clearly that this data has to be public,” says Wüest. “To my knowledge, the government of Rwanda never lived up to that.” Mudakikwa says that data relating to the gas-extraction companies are confidential, but lake-profile data can be obtained if researchers write a letter to the director-general of REMA explaining what they need and why they need it.

The monitoring programme only recently moved under the remit of REMA; until April, it was under the Rwanda Energy Group, which is also the country’s national energy utility company. The programme’s new website hasn’t yet been set up. The authority is currently revising the MPs, Mudakikwa says, in part to better outline its data-sharing policies.

Augusta Umutoni, who headed the LKMP until this April, says she is proud of the technical team she helped to set up, and thinks the Rwandan government is committed to keeping the monitoring effort going. But, she adds, governments sometimes find their budgets stretched thin, or become bogged down in bureaucracy. “The governments and operators will have to

work together,” she emphasizes. The MPs also recommended the creation of a bilateral regulatory authority shared by Rwanda and the DRC; this has not yet happened, confirms Mudakikwa.

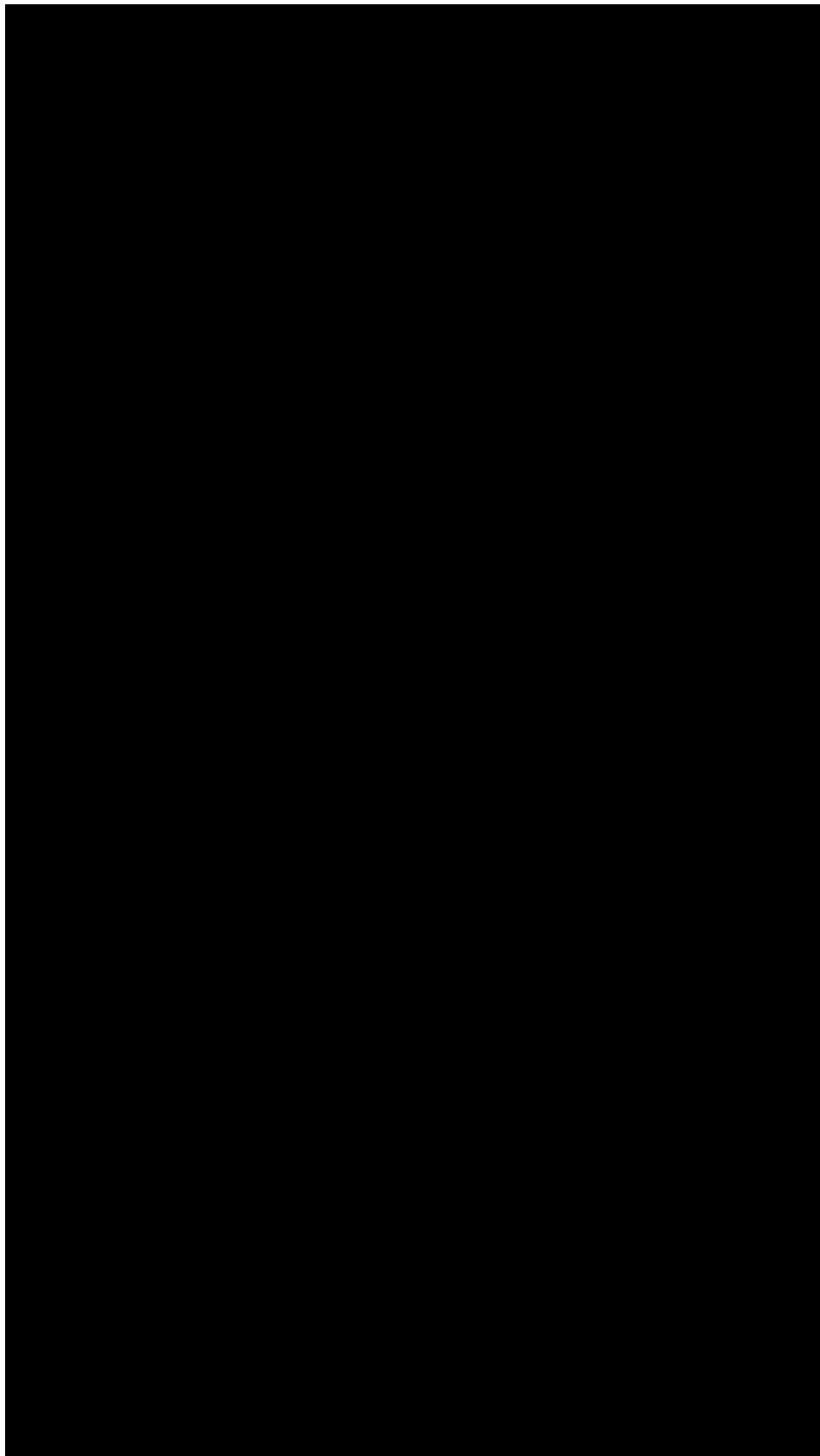
The combination of Lake Kivu’s monetary value, its potential explosive capacity, and the huge range of opinions about how to best deal with it, makes emotions run high among the scientists who work there. “It has become an obsession for me to understand what’s going on in this lake,” says Hirslund. “When you start working with Lake Kivu, you get passionate,” agrees Umutoni.

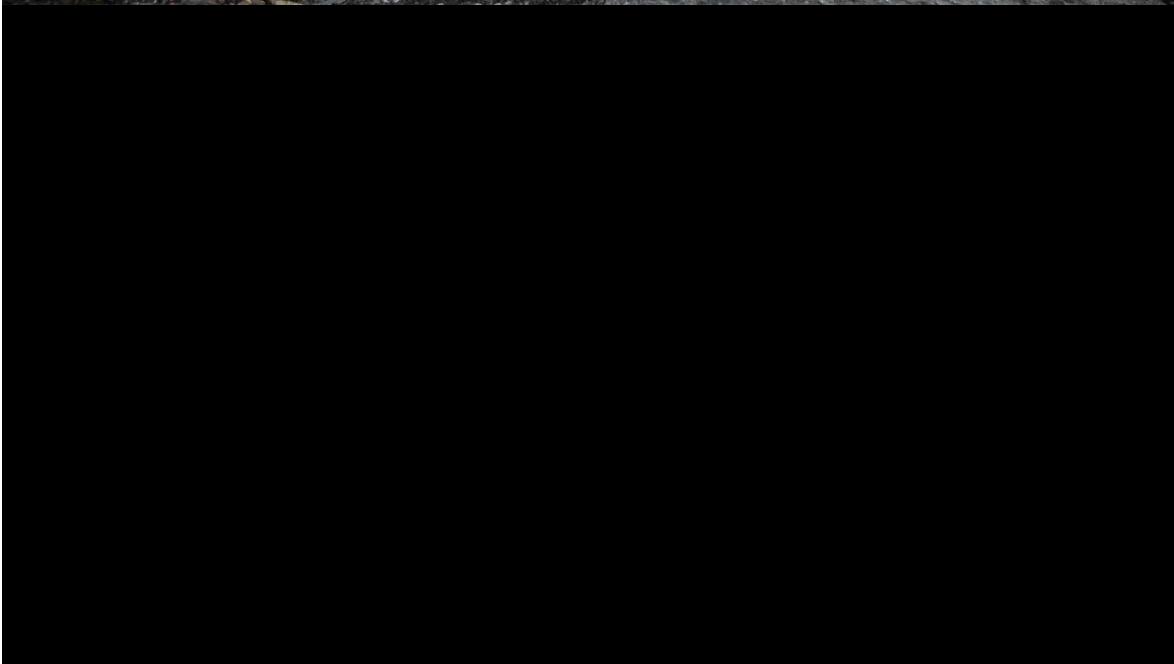
Taking gas out of the lake should be making it safer, says Mudakikwa, but there are some things — such as a volcanic eruption — that no scientist, company or regulatory authority can counter or prevent. “If it’s Mother Nature, you can’t fight Mother Nature.”

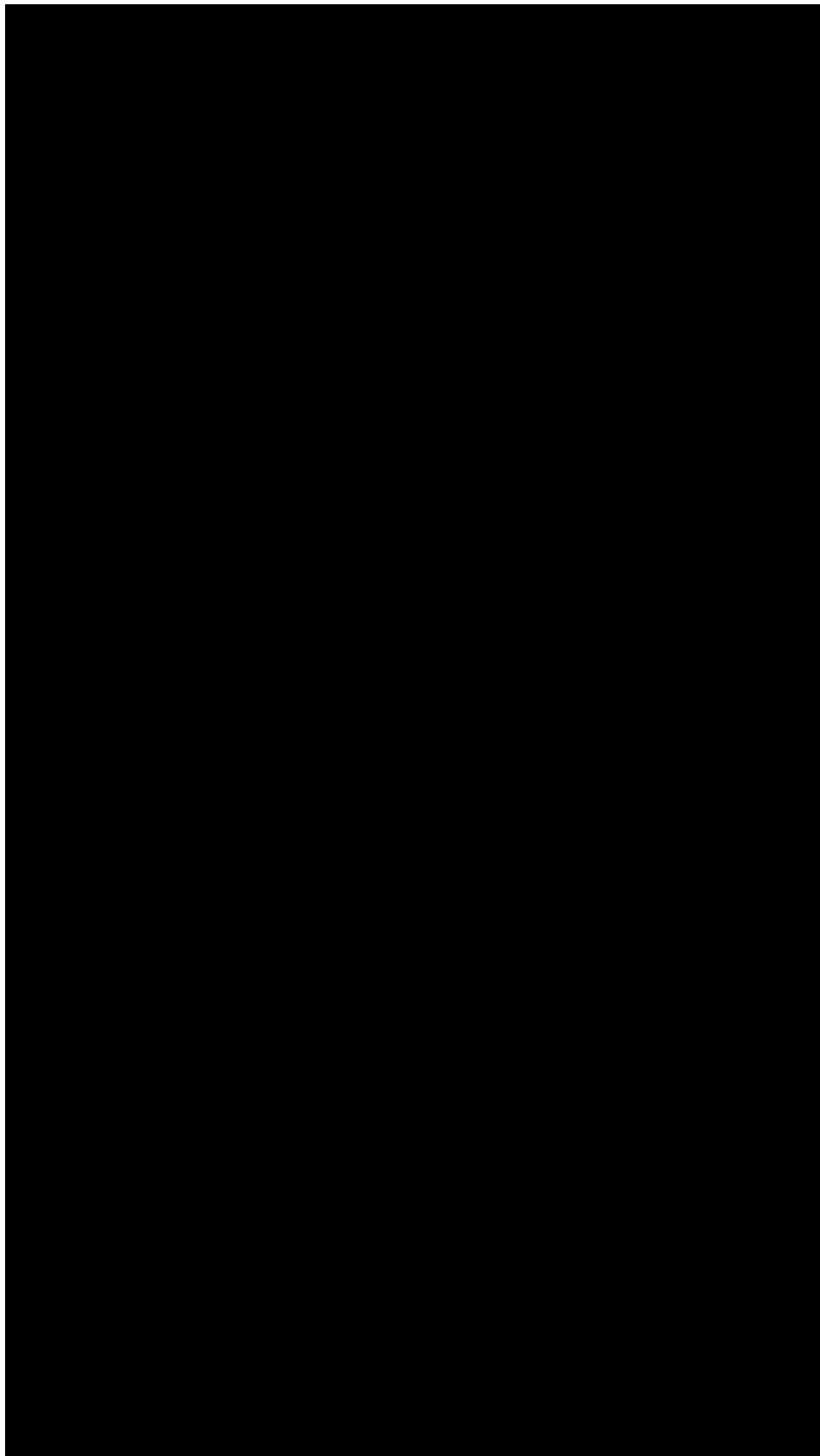
---

**Nicola Jones** is a science journalist based in Pemberton, Canada.











Credit: Guerchom Ndebo for *Nature*

---

## References

1. Halbwachs, M. Sabroux, J.-C. & Kayser, G. *J. Afr. Earth Sci.* **167**, 103575 (2020).  
[Article](#) [Google Scholar](#)
2. Votava, J. E., Johnson, T. C. & Hecky R. E. *Proc. Natl Acad. Sci. USA* **114**, 251–256 (2017).  
[PubMed](#) [Article](#) [Google Scholar](#)
3. Hirslund, F. *J. Afr. Earth Sci.* **161**, 103614 (2019).  
[Article](#) [Google Scholar](#)
4. Schmid, M., Halbwachs, M., Wehrli, B. & Wüest, A. *Geochem. Geophys. Geosys.* **6**, Q07009 (2005).

[Article](#) [Google Scholar](#)

5. Bärenbold, F. *et al.* *PLoS ONE* **15**, e0237836 (2020).

[PubMed](#) [Article](#) [Google Scholar](#)

6. Hirslund, F. & Morkel, P. *J. Afr. Earth Sci.* **161**, 103672 (2020).

[Article](#) [Google Scholar](#)

---

**SPRINGER NATURE**

Springer Nature

- [Privacy Policy](#)
- [Use of cookies](#)
- [Legal notice](#)
- [Terms & Conditions](#)
- [Accessibility statement](#)

© 2021 Springer Nature Limited. All rights reserved.

[Top](#)



## [Built with Shorthand](#)

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02523-5>

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

## Books & Arts

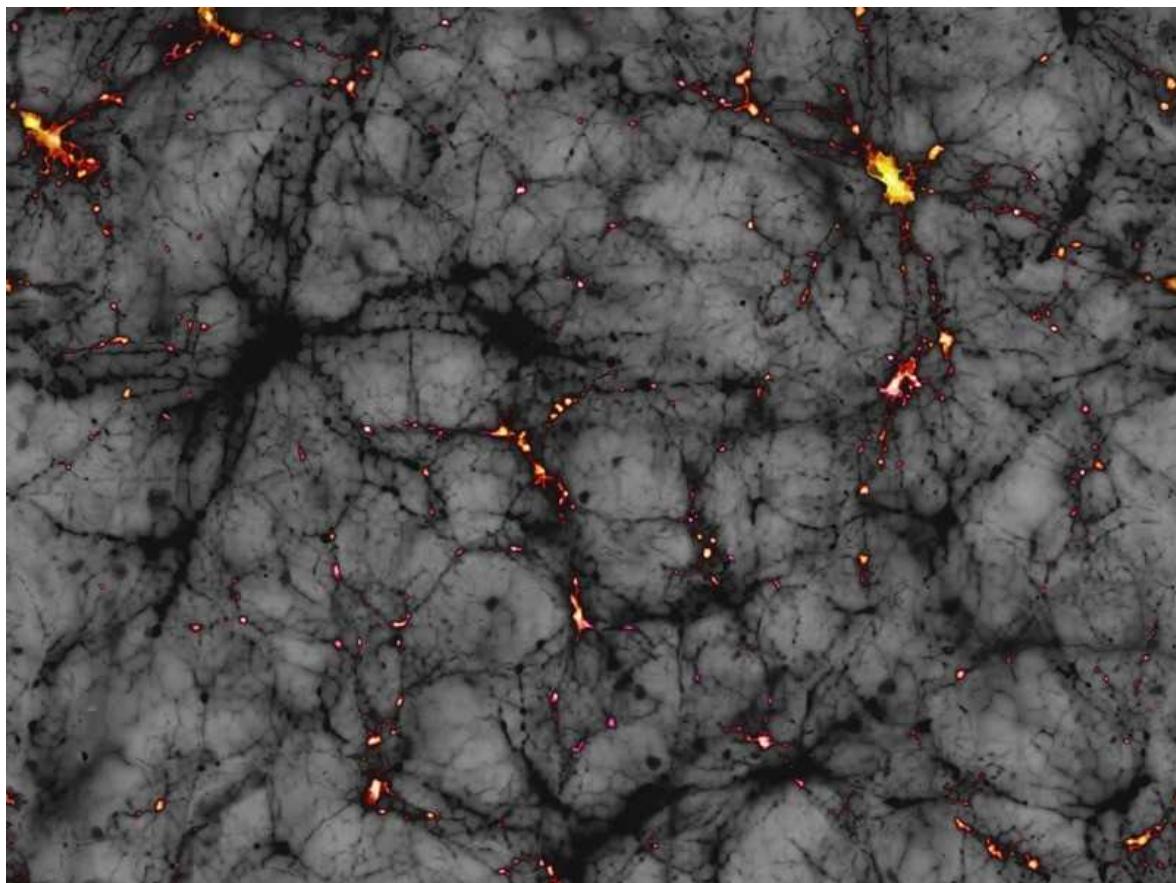
- **Who is allowed to have wild ideas in physics?** [ 20 September 2021]  
Book Review • A cosmologist reflects on barriers to diversity of thought in dark-matter and dark-energy research.
- **A meander around many circulatory systems** [ 20 September 2021]  
Book Review • Of hearts, and myriad other ways natural selection has hit on to sustain multicellular life.
- **Intellectual bees, doyenne of dark matter, and mathematical grief: Books in brief** [ 06 September 2021]  
Book Review • Andrew Robinson reviews five of the week's best science picks.

- BOOK REVIEW
- 20 September 2021

# Who is allowed to have wild ideas in physics?

A cosmologist reflects on barriers to diversity of thought in dark-matter and dark-energy research.

- [Anil Ananthaswamy](#) ✉



The presence of dark matter (light spots, illustration) is inferred from its effects on visible matter. Credit: Mark Garlick/SPL

## **Fear of a Black Universe: An Outsider's Guide to the Future of Physics**

*Stephon Alexander Basic (2021)*

Dark matter and dark energy are mysterious components of the cosmos that have thrown monkey wrenches into our understanding. In *Fear of a Black Universe*, Stephon Alexander writes of efforts to make sense of this “dark sector”. They have stalled, he posits, partly because researchers are wary of bold explanations for these unknowns. He offers a personal exploration of whether science is receptive to ideas that violate norms and expectations. And Alexander, a Black theoretical physicist, asks: is the search for answers in modern physics hindered by an establishment afraid to entertain the ideas of those it considers outsiders?

A physicist at Brown University in Providence, Rhode Island, Alexander had an epiphany while hiking in Trinidad and Tobago, his birthplace. Although brimming with ideas, he realized that he had been dodging research on seemingly controversial topics because of the risks to his professional standing. “Black persons in scientific circles are often met with skepticism about their intellectual capabilities,” he writes. “My exploratory, personal style of theorizing, when coupled with my race, often creates situations where my white colleagues become suspicious and devalue my speculations.” This book — his second — makes a poignant case for why everyone deserves equal opportunities to let their imagination soar.



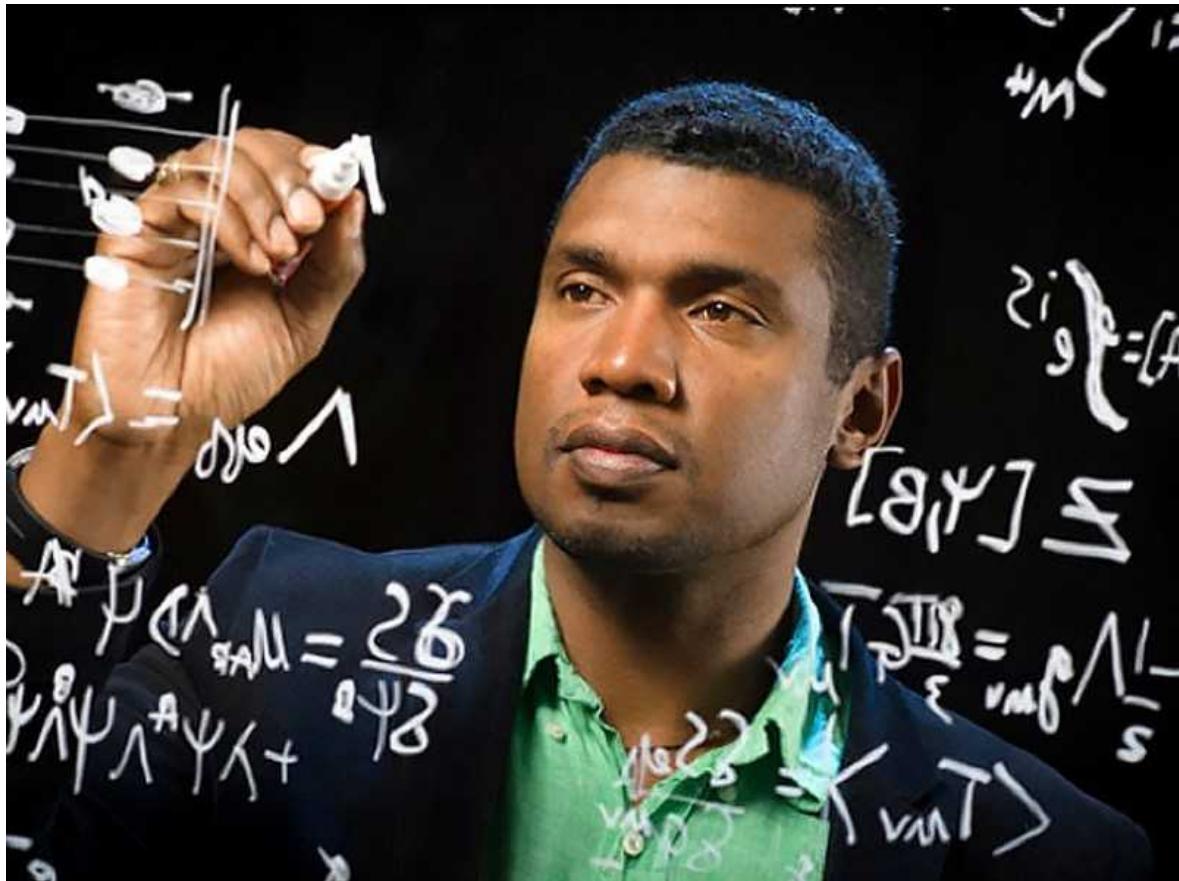
## Vera Rubin, astronomer extraordinaire — a new biography

Alexander discusses three fundamental principles of physics. The first is that of invariance — the laws of physics are unchanging for observers moving relative to each other at constant speeds. This principle underlies Isaac Newton's laws of mechanics, and similar ideas allowed Albert Einstein to develop his special and general theories of relativity. Second, he focuses on superposition: the state of a quantum mechanical system is expressed as a combination of all the possible states that the system can be in. Third is the principle of emergence: "Systems with interacting elementary constituents can exhibit novel properties that are not possessed by the constituents themselves." For example, superconductivity, the phenomenon in which electrical resistance disappears in certain materials under certain conditions, is an emergent property of some quantum-mechanical systems.

## Riffs on the classics

This is a fresh way of introducing some basics, but Alexander's brisk, brief forays demand a lot of his readers. An accomplished jazz musician, Alexander admits that his narrative and structure will proceed like an improvisation (his first book was *The Jazz of Physics* in 2016). Much as an untrained ear can find it hard to appreciate a complex solo, those unschooled in relativity, quantum mechanics, string theory and quantum gravity might struggle to keep up.

Alexander riffs through a host of grand ideas: the nature of the Big Bang and the questions it begets; the origin of life; the role of consciousness in quantum mechanics and the evolution of the Universe; theories that seek to reconcile general relativity with quantum mechanics; and more. Others have written tomes on each of these topics, and Alexander has no doubt thought deeply about them, but they're hard to corral cogently into a couple of hundred pages of non-linear narrative.



Cosmologist Stefon Alexander also explored the importance of improvisation in his first book, *The Jazz of Physics*. Credit: John Sherman

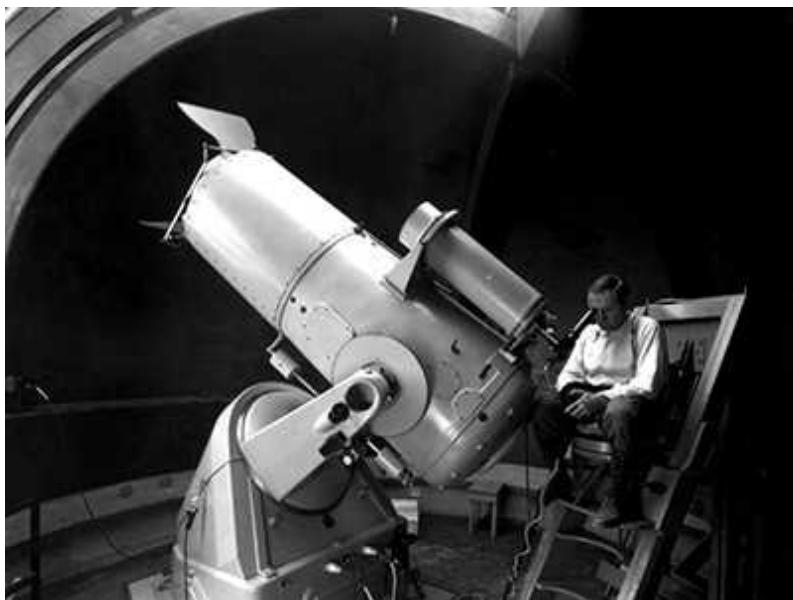
Take the chapter ‘Dark ideas on alien life’. It’s an account of a wild thought experiment that Alexander dreamed up with his friend Jaron Lanier, a virtual-reality pioneer. What if, they ask, there are numerous alien civilizations running powerful quantum computers that tie topological knots in the fabric of space-time to do computations, using gravitational-wave detectors to read from and write information to the vacuum of space-time? It is speculation piled upon speculation — breathless stuff.

Alexander writes that this “bizarre notion” could explain why the observed amount of dark energy in the Universe is nearly 120 orders of magnitude smaller than expected from theoretical considerations: maybe “the aliens used dark energy as a resource to run their ultimate computers in much the same way we devour oil to run our cars and jets”. Why would they? To enjoy high-quality virtual reality, of course. Alexander’s leaps of

imagination follow the strong tradition of thought experiments in physics, but their import might be accessible only to cognoscenti.

## Case studies in diversity

Along the way, Alexander wants to convince the reader that the lack of diversity in science diminishes the quality of the research accomplished, as well as being a social-justice concern. Two of his stories exemplify the issues. One is about James Gates, an African American theorist, whose work on supersymmetry (an extension to the standard model of particle physics) in the 1990s with Hitoshi Nishino got little attention. According to Alexander, similar work more than a decade later, called the ABJM theory (after the last names of the researchers who developed it: Ofer Aharony, Oren Bergman, Daniel Jafferis and Juan Maldacena), was hailed as a landmark result. Alexander challenges us to reflect on why few researchers (himself included) noticed Gates and Nishino's earlier work.



### [Deciphering dark matter: the remarkable life of Fritz Zwicky](#)

The other story is more personal. Alexander gives a harrowing account of being told by a white visiting colleague and friend that his fellow postdocs at Stanford University in California might be shunning him because “they feel that they had to work so hard to get to the top and you [Alexander] got in

easily, through affirmative action”. It goes without saying that they had no idea how hard Alexander had had to work, or what barriers he’d had to vault, to make it in physics, coming from a poor family growing up in the Bronx, New York City, in the 1980s.

Stung, and realizing that he had to showcase his strengths independently, Alexander began working alone in a café, without help from colleagues. Here, he honed the outside-the-box thinking that led to a paper (with his mentor Michael Peskin) providing a new theory for why there is much more matter than antimatter in the Universe ([S. H. S. Alexander et al. Phys. Rev. Lett. 96, 081301; 2006](#)). His account offers powerful insight into the systemic forces that work against inclusion.

There’s no doubt that physics has a diversity problem in the United States — one of the biggest in all the sciences. According to the American Institute of Physics, in 2012, Black or African Americans, who comprise about 13.4% of the US population, made up only 2.1% of physics faculty members. In 2018, members of the American Physical Society’s inclusion team warned that although about one-third of university-age US citizens are African Americans, Hispanic Americans or Native Americans, less than 11% of bachelor’s degrees in physics are awarded to people from these groups. The figure is just 7% for PhDs — around 60–70 students each year (see [T. Hodapp and E. Brown Nature 557, 629–632; 2018](#)).

In addition to the impact of historical and structural racism on the gatekeeping of ideas, other sociological factors advantage some avenues of research over others. The community of string theorists is large and well funded and can out-compete other theories of quantum gravity, for instance — as is explored in books such as *The Trouble With Physics* (2006) by Lee Smolin. And many ideas are discarded simply because they are bad. *Fear of a Black Universe* might have been richer for a more searching look at the way these factors interact. Nevertheless, it’s a timely reminder of the need to hear a wider variety of voices in physics, as in all the sciences.

*Nature* **597**, 471–472 (2021)

doi: <https://doi.org/10.1038/d41586-021-02526-2>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02526-2>

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

- BOOK REVIEW
- 20 September 2021

# A meander around many circulatory systems

Of hearts, and myriad other ways natural selection has hit on to sustain multicellular life.

- [Henry Nicholls](#) 0



Horseshoe crabs' blood is blue: their oxygen-transport protein is copper-based. Credit: Getty

## Pump: A Natural History of the Heart Bill Schutt Algonquin (2021)

Rich in meaning and metaphor, the word ‘heart’ conjures up many images: a pump, courage, kindness, love, a suit in a deck of cards, a shape or the most important part of an object or matter. These days, it also brings to mind the global increase in heart attacks and cardiovascular damage that attends COVID-19. As a subject for a book, the heart is an organ with a lot going for it.

Enter zoologist Bill Schutt. His book *Pump* refuses to tie the heart off from the circulatory system, and instead uses it to explore how multicellular organisms have found various ways to solve the same fundamental challenge: satisfying the metabolic needs of cells that are beyond the reach of simple diffusion. He writes of the co-evolution of the circulatory and respiratory systems: “They cooperate, they depend on each other, and they are basically useless by themselves.”

At his best, Schutt guides us on a journey from the origin of the first contractile cells more than 500 million years ago to the emergence of vertebrates, not long afterwards. He takes in, for example, horseshoe crabs, their blood coloured blue by the presence of the copper-based oxygen-transport protein haemocyanin (equivalent to humans’ iron-based haemoglobin).



## Animal crackers

We learn that insects, lacking a true heart, have a muscular dorsal vessel that bathes their tissues in blood-like haemolymph. Earthworms, too, are heartless but with a more complex arrangement of five pairs of contractile vessels. Squid and other cephalopods have three distinct hearts.

There are plenty of zoological nuggets to enjoy along the way. The tubular heart of a sea squirt, for instance, contains pacemaker-like cells that enable it to pump in one direction and then the other. Some creatures need masses of oxygen, others little, leading to more diversity. The plethodontids (a group of salamanders) have neither lungs nor gills, he explains: their relatively small oxygen requirements are met by diffusion through the skin.

## Cardiac records

Hagfish can get by with the lowest recorded aortic pressure of any vertebrate, between 5.8 and 9.8 mm Hg. A giraffe's heart, by contrast, must generate extraordinary pressures — up to 280/180 mm Hg — to send blood up its 2-metre-long neck to its brain. Hummingbird hearts can reach an astonishing 1,260 beats per minute. Shrew hearts must work faster still, each cardiac cycle lasting just 43 milliseconds — a heart rate that must be “awful damn close” to the electrophysiological maximum.

Schutt refers to one of his own research interests, cold adaptation in bats: a physiological trick that sees the heart rate collapse from well over 500 beats per minute during flight to less than 20 beats per minute during hibernation. These metabolic extremes might help to explain why bats are a reservoir for so many viruses ([A. T. Irving et al. \*Nature\* 589, 363–370; 2021](#)). However, *Pump* contains no reference to SARS-CoV-2 and the many ways — direct and indirect — in which this particular coronavirus seems to affect the cardiovascular system ([M. Nishiga et al. \*Nature Rev. Cardiol.\* 17, 543–558; 2020](#)).



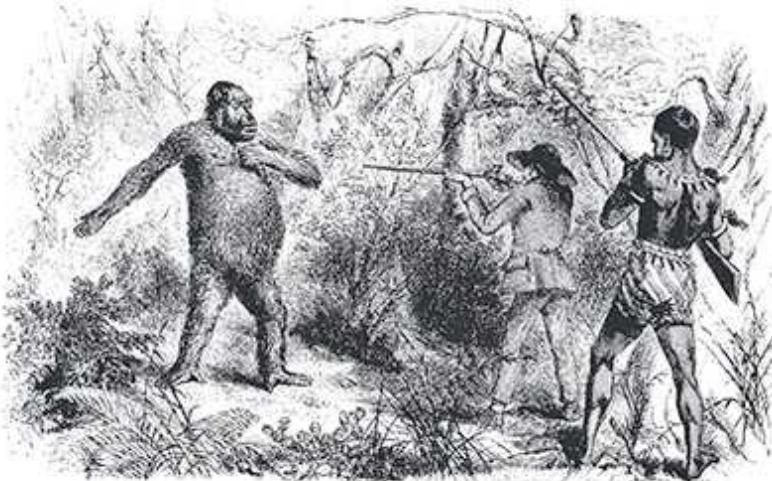
A shrew's heartbeat lasts for just 43 milliseconds. Credit: Getty

As Schutt works his way around the evolutionary tree, he is keen to stress that “there should be no bragging rights associated with the fact that some circulatory systems are quite complex while others are relatively simple. The key here is that all of them work.” Rather than seeing humans as the highest peak on the evolutionary landscape, as writers might have in previous centuries, he celebrates the functional equivalence of non-human circulatory systems, successful arrangements that have propelled their owners to other summits. Many of these peculiar anatomies are brought to life in beautiful line drawings by award-winning illustrator Patricia Wynne.

## Conversation piece

Schutt's try-hard tone will not be for everyone. “Hey, guys, don't take this the wrong way,” he writes as part of a dialogue with organisms (such as flatworms) that have no “circulatory-system junk”. “We interrupt this

chapter . . .” he announces, to suggest that anyone without a carbon monoxide detector should “go purchase one. I’ll wait.” Perhaps the most surprising sentence to have made the cut, despite what Schutt acknowledges as “significant and outstanding editorial input”, comes after one of many digressions: “For some reason, I thought this information was worth including here.”



### Beastly surprises

As Schutt turns from comparative anatomy to historical interpretations of the heart, and then to more recent milestones in cardiovascular medicine, the off-piste jaunt loses its way. From the ancient Egyptian belief that the heart held a record of a person’s good and bad deeds to the use of stem cells and 3D printers to build new organs, there is a lot of ground to cover. Highlights include the contributions of thirteenth-century Syrian polymath Ibn al-Nafis, along with Spanish theologian Michael Servetus and Italian anatomist Matteo Realdo Colombo in the sixteenth century, to our understanding of the true relationship between the heart and lungs. The English physician William Harvey (1578–1657) is most commonly associated with this discovery.

### **Priority shift**

Other choices are puzzling. A chapter dedicated to Charles Darwin's long illness after his voyage on *The Beagle* is a case in point. Although the mystery involves the naturalist's heart and is undeniably interesting, why give this space while a landmark in cardiovascular medicine such as the invention of the heart–lung machine is barely mentioned? Schutt gives scant attention, either, to how epidemiology, electrocardiography, bypass surgery and stents have transformed the diagnosis and treatment of cardiovascular disease. All have contributed to a dramatic decline in cardiovascular mortality by around 60% over the past 50 years. As for the legacy of the damage wrought to hearts — in every sense — by the COVID-19 pandemic? That's a book that needs to be written.

*Nature* **597**, 472–474 (2021)

doi: <https://doi.org/10.1038/d41586-021-02524-4>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02524-4>

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

- BOOK REVIEW
- 06 September 2021

# Intellectual bees, doyenne of dark matter, and mathematical grief: Books in brief

Andrew Robinson reviews five of the week's best science picks.

- [Andrew Robinson](#) 0



# Silent Earth

AVERTING THE INSECT APOCALYPSE

DAVE GOULSON

**Silent Earth**

Dave Goulson *Jonathan Cape* (2021)

Biologist Dave Goulson loves insects. As a child, he fed yellow-and-black caterpillars and watched them become cinnabar moths. As an adult, he showed how bumblebees avoid wasting time on a flower visited by another bee — by sniffing it for the fresh whiff of smelly feet. They also detect a decrease in the electrostatic charge on its pollen. Bees are “the intellectual giants of the insect world”, he writes enchantingly, while pondering an alarming estimated 75% decline in global insect populations over half a century.



# Bright Galaxies, Dark Matter, and Beyond

*The Life  
of Astronomer  
Vera Rubin*

Ashley Jean Yeager



Bright Galaxies, Dark Matter, and Beyond

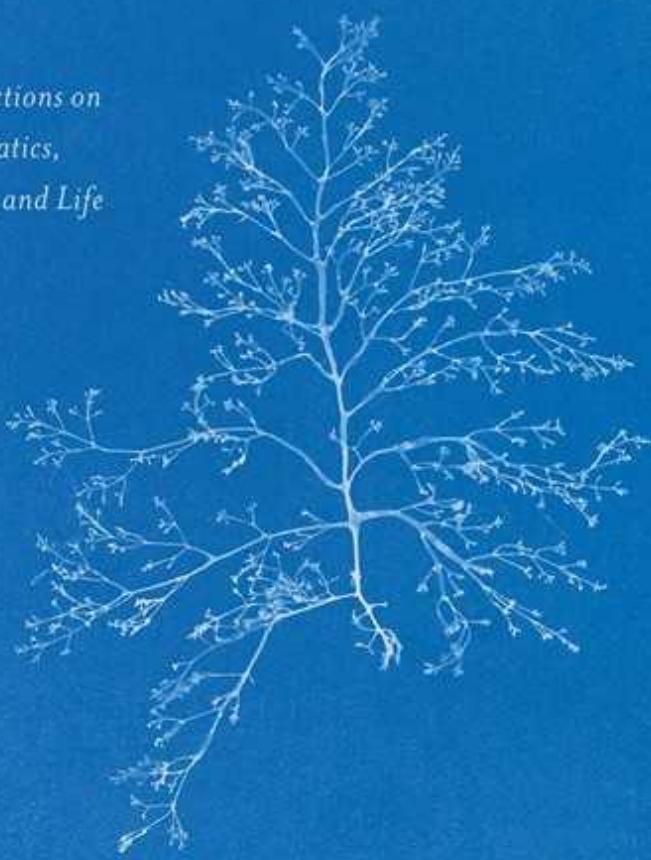
Ashley Jean Yeager *MIT Press* (2021)

‘More matter than meets the eye’ is a chapter title of this insightful biography of the pioneering astronomer Vera Rubin by science journalist Ashley Yeager, who interviewed her in later life. Best known for her observations of galactic rotation rates, which provided evidence for the existence of dark matter, Rubin also campaigned for equality in science. Her many honours did not include a Nobel prize, but a new observatory in Chile bears her name and this is the second biography of her in a year (see [A. Abbott \*Nature\* 591, 523–524; 2021](#)).

G E O M E T R Y

*of G R I E F*

*Reflections on  
Mathematics,  
Loss, and Life*

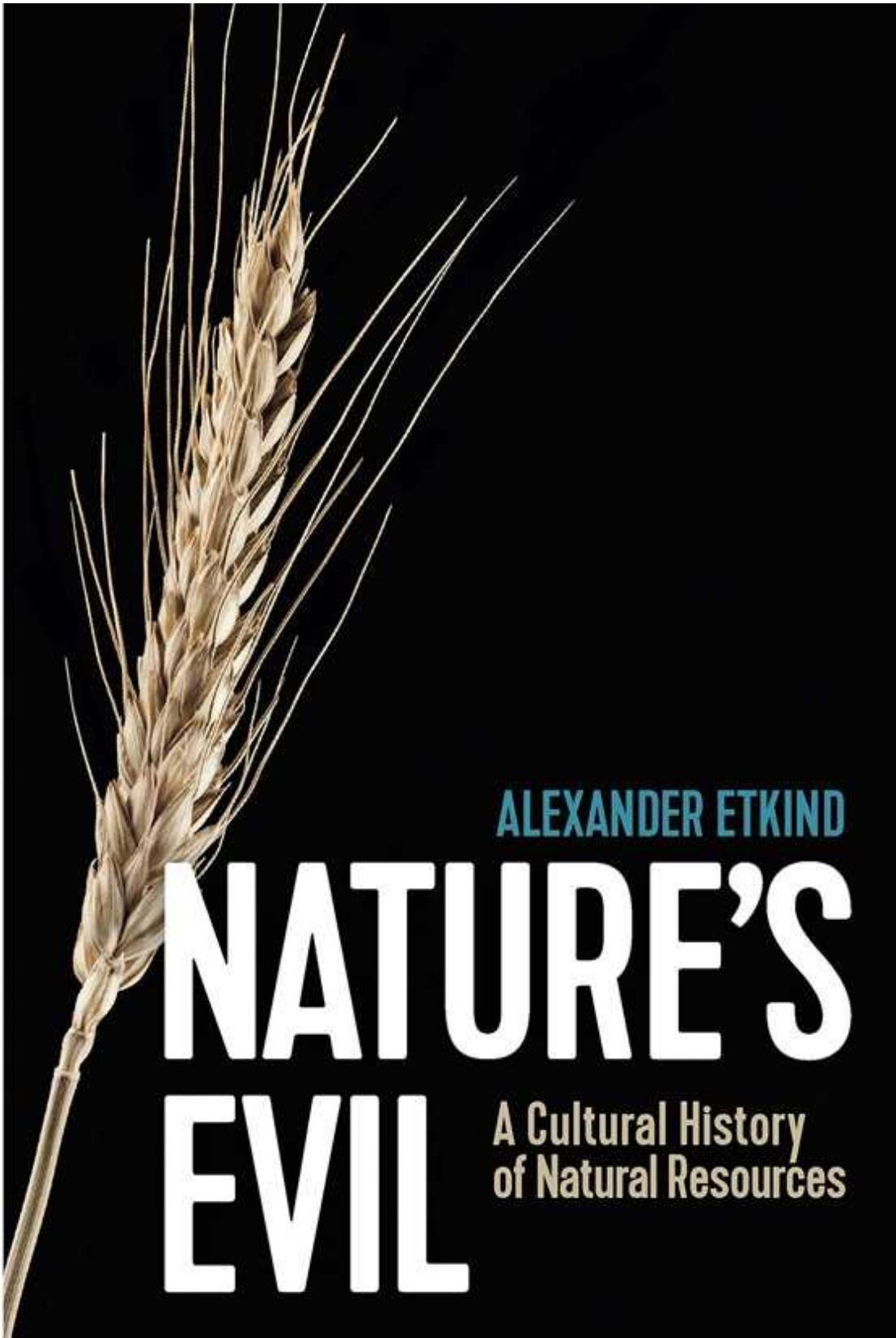


M I C H A E L F R A M E

**Geometry of Grief**

Michael Frame *Univ. Chicago Press* (2021)

This brief, intriguing personal meditation is inspired by mathematician Michael Frame's lifelong love of geometry — including 20 years' collaboration with fractal geometer Benoit Mandelbrot — and the childhood loss of his aunt, who set him on his career path. He writes: "Grief informs geometry and geometry informs grief." How so? His epiphany on first understanding any beautiful mathematical idea is always tinged with sadness, because it is unrepeatable. With quirky illustrations, he integrates the lives of his Mom and Dad.



Nature's Evil

## Alexander Etkind *Polity* (2021)

In detailed chapters on grain, animal products, sugar, hemp, metals, peat, coal and oil, historian Alexander Etkind explores how nature and its commodification has shaped states and societies, as the pursuit of power and wealth has degraded people and despoiled the planet. His Eurocentric survey weaves together material, intellectual, economic, ecological and moral history to reflect on “the mess we have made of our world”. To predict the outcomes of our choices, he argues, it pays to know the consequences of choices that people made in the past.

---

---

THE COLLECTED PAPERS OF

---

# Albert Einstein

---

---

VOLUME 16

THE BERLIN YEARS:  
WRITINGS & CORRESPONDENCE,  
JUNE 1927–MAY 1929

ENGLISH  
TRANSLATION

Diana Kormos Buchwald, Ze'ev Rosenkranz, József Illy, Daniel J. Kennefick,  
A. J. Kox, Dennis Lehmkuhl, Tilman Sauer, AND Jennifer Nollar James  
EDITORS

Jennifer Nollar James, William D. Brewer, AND Steven Rendall  
TRANSLATORS

---

**The Collected Papers of Albert Einstein, Volume 16**

Eds Diana Kormos Buchwald *et al.* Princeton Univ. Press (2021)

Albert Einstein's collected papers began publication in 1987. The 16th of these uniquely comprehensive volumes covers 1927–29, up to Einstein's 50th birthday, when he hid from acclaim. It includes the 1927 Solvay Conference on quantum mechanics, where he sparred with Niels Bohr but scribbled a note: "Who knows who'll be laughing in a few years?" He also engaged further in politics, dubbing himself an anti-fascist, and hired assistant Helen Dukas, who preserved his letters post-mortem, creating his vast archive.

*Nature* **597**, 473 (2021)

doi: <https://doi.org/10.1038/d41586-021-02403-y>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02403-y>.

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

# Opinion

- **Funders: cover APCs for African scholars — and do more**  
[ 21 September 2021]  
Correspondence •
- **Afghanistan: Taliban's return imperils maternal health** [ 21  
September 2021]  
Correspondence •
- **Changing the wrapping won't fix genetic-racism package**  
[ 21 September 2021]  
Correspondence •

- CORRESPONDENCE
- 21 September 2021

# Funders: cover APCs for African scholars — and do more

- [Bård Vegar Solhjell](#)<sup>0</sup>,
- [Kjersti Thorkildsen](#)<sup>1</sup> &
- [Grete Benjaminsen](#)<sup>2</sup>

Although one of the principles of Plan S is that open-access journals must waive article-processing charges (APCs) for authors from low-income countries, that does not always happen, as Addisu Mekonnen *et al.* point out ([Nature 596, 189; 2021](#)). So, the Norwegian Agency for Development Cooperation (Norad) — which has long encouraged open-access publishing when funding research and higher education in the global south — covers APCs in its projects. This alone is not, however, a sustainable solution.

Regional and national research councils and international donors should also invest in African open repositories and local grant schemes to cover APCs, as well as in more open journals and publishing platforms of high quality. Such approaches benefit all scholars, especially those from low-income countries. Currently, almost 12,000 journals that are free to publish in and free to read are registered in the Directory of Open Access Journals. More are needed.

Norad supports several digital public-goods initiatives with open platforms and open content. One such is the open-source District Health Information Software 2 (DHIS-2). This is the world's largest health-management information system, in use by 73 low- and middle-income countries.

*Nature* **597**, 475 (2021)

*doi: <https://doi.org/10.1038/d41586-021-02552-0>*

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02552-0>

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

- CORRESPONDENCE
- 21 September 2021
- Update [24 September 2021](#)

# Afghanistan: Taliban's return imperils maternal health

- [Shohra Qaderi](#)<sup>0</sup>,
- [Attaullah Ahmadi](#)<sup>1</sup> &
- [Don Eliseo Lucero-Prisno III](#)<sup>2</sup>

Afghanistan has one of the highest maternal death rates in the world, despite gains in women's health over the past 20 years (see [go.nature.com/39burgc](http://go.nature.com/39burgc)). Experience suggests that the Taliban's takeover of the country will further imperil mothers' health and well-being.

During the previous reign of the Taliban (1996–2001), the maternal and neonatal death rates worsened as a consequence of the complex synergy of social, demographic, medical, economic and cultural factors ([S. A. M. Najafizada et al. Cent. Asian J. Glob. Health 6, 240; 2017](#)). Restrictions to women's lives included allowing only female health workers to examine them, limited access to quality health services — particularly obstetric care — and minimal opportunities for education and work. These increased the risk of giving birth at home with no prenatal or natal care ([C. Palmer Lancet 352, 734; 1998](#)).

The United Nations 2030 goals for sustainable development include reduction of global maternal mortality to less than 70 deaths per 100,000 live births. Afghanistan's latest figure of 638 per 100,000 is now more likely to grow than to shrink. In our view, rectifying this should be an international priority.

*doi: <https://doi.org/10.1038/d41586-021-02551-1>*

## Updates & Corrections

- **Update 24 September 2021:** The affiliation for Shohra Qaderi has been updated.

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02551-1>

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

- CORRESPONDENCE
- 21 September 2021

# Changing the wrapping won't fix genetic-racism package

- [Latifa Jackson](#) ORCID: <http://orcid.org/0000-0002-0949-978X><sup>0</sup>,
- [Krystal S. Tsosie](#) ORCID: <http://orcid.org/0000-0002-7291-670X><sup>1</sup> &
- [Keolu Fox](#) ORCID: <http://orcid.org/0000-0003-4215-5273><sup>2</sup>

*Nature* misses a chance to grant agency to marginalized communities in inviting Alice Popejoy to point out that altering racial classifications will not absolve power imbalances in genetics ([Nature 596, 463; 2021](#)).

In 1785, philosopher Christoph Meiners reduced continental-scale diversity to an imperial classification system to subjugate colonized peoples. This system is still used by geneticists, and lingers beyond terms such as Caucasian. New ethnonyms replaced older terms ('mongoloid' became 'Asian', for instance) but failed to redress underlying racism. And socially constructed categories are used in biologically essentialist 'race correction' to model disease risks (see, for example, [D. E. Roberts Lancet 397, 17–18; 2021](#)).

To demolish genetic racism, geneticists must defer to communities to self-define their 'belongingness' (see, for example, [K. S. Tsosie Curr. Opin. Genet. Dev. 62, 91–96; 2020](#)). Any unequal system of classification that reifies race, ethnicity and ancestry for biological insight reproduces the obstacles it attempts to dismantle and does not solve the causes of health disparities.

We advocate empowering communities to label themselves; to undertake ethnographies to contextualize research findings; and to self-determine research they deem beneficial.

*Nature* **597**, 475 (2021)

doi: <https://doi.org/10.1038/d41586-021-02553-z>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02553-z>

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

# Work

- **Discovering allyship at a historically Black university** [ 16 September 2021]  
Career Column • Adrienne Nugent's postdoctoral programme at Hampton University showed her what it felt like to be a member of a minority group on a committee.
- **Cash boost looms for historically Black US colleges and universities** [ 20 September 2021]  
Career News • Legal wins and federal budget proposals could address years of underfunding.
- **Single-cell proteomics takes centre stage** [ 20 September 2021]  
Technology Feature • Deducing the full protein complement of individual cells has long played second fiddle to transcriptomics. That's about to change.
- **Preserving pieces of history in eggshells and birds' nests**  
[ 20 September 2021]  
Where I Work • Museum curator Douglas Russell catalogues and maintains specimens that offer a glimpse into the breeding behaviours of birds both living and extinct.

- CAREER COLUMN
- 16 September 2021

# Discovering allyship at a historically Black university

Adrienne Nugent's postdoctoral programme at Hampton University showed her what it felt like to be a member of a minority group on a committee.

- [Adrienne Nugent](#) ✉



Students on the campus of Hampton University, a historically Black institution. Credit: Hampton University

I have always considered myself an ally to people from marginalized groups. But my two-year postdoctoral programme in the cancer research centre at Hampton University in Virginia, one of more than 100 historically Black colleges and universities (HBCUs) in the United States, gave me an entirely new understanding of allyship. The experience was transformative. As a white person who grew up in a small town in New England with a homogeneous population, I was totally unprepared for how I would learn and develop at an HBCU.

In the first of many lessons in allyship, I realized I hadn't heard the term 'predominantly white institution' (PWI) until I arrived at Hampton in February 2017, despite having studied at Duke University, a PWI in Durham, North Carolina, and having completed a postdoctoral programme at the US National Institutes of Health (NIH) in Bethesda, Maryland.

I started at Hampton after my husband, who was in the military, was posted at the nearby Langley Air Force Base. On my first day, I passed hundreds of people, none of whom looked like me. Over the next two years, I started to understand some of the issues that scientists who are part of minority-ethnic populations face daily. Although my experience provided a glimpse of what it is like to be a student and researcher at a minority-serving institution, I acknowledge that I as a white person can never fully understand my colleagues' perspectives. But from their frank and thoughtful advice, I have learnt some things that might be useful to others in a similar position to mine.

I benefited from teaching and mentoring university students who were vocal and upfront about their struggles and solutions, and who candidly told me what being an ally means. In many ways, they mentored me. Hearing their stories first-hand made me reflect on what it must feel like to be the only Black person in a graduate-school cohort or on a grant committee, facing the expectation that you will represent your whole population or culture. I had heard anecdotes of principal investigators at HBCUs who'd sat on review committees on which they were the only person of colour out of 30 people reviewing grants for minority-serving institutions. That pressure must be intense, and the margin for failure, non-existent.

At most major US research institutions, it seems as though seminars occur hourly, everybody has an NIH research grant and samples are easily accessible at the university hospital. Researchers at these places are part of an institutional cycle of winning grants and creating networks, and getting access to published studies is not a problem.

HBCUs, by contrast, are mostly smaller liberal-arts schools, and very few have major medical centres. But there, too, networking is crucial. In fact, networks at HBCUs are even more important than they might be at PWIs, because of differences in available resources. For example, HBCUs often lack funding to support access to most scientific literature, so researchers can find themselves spending lots of time, effort and mental energy trying to find and access articles rather than focusing on research. Yet despite these systemic barriers, the Hampton community had an incredible atmosphere and an unmistakable drive to persevere and overcome these hurdles.

One day, I asked my Hampton adviser, Luisel Ricks-Santi, a cancer geneticist, about attending an NIH conference on research that interested me. She suggested that perhaps the NIH should come to Hampton instead. That didn't happen, but the contrast it highlighted was enlightening. The culture of the scientific enterprise includes the unspoken expectation that 'everybody should come to us and learn from us'. In fact, the scientific community would benefit if people with PhDs and MDs at major institutions went to HBCUs to learn from their communities. But there would have to be a real interest in learning and not just imposing fancy new techniques, offering second authorship in collaborations or throwing money around (although more money would help).

Researchers at HBCUs need to lead collaboration conversations from the initial stages of study design to data collection and analysis. The more voices we can incorporate into these discussions, the better the scientific and medical communities, and the people they serve, will be. HBCUs have been leaders in scholarship on health disparities and community relations for almost 200 years, and it's important that scientists recognize their role and empower HBCUs to guide us in our collective effort to achieve health equity.

I now work at Invitae, a genetic-health company in San Francisco, California, and have connected the company and Hampton to address inequalities in genetics. Together, we're building a career pipeline, mentoring platform and educational series to increase diversity in the science and biotech workforce.

*Nature* **597**, 577-578 (2021)

doi: <https://doi.org/10.1038/d41586-021-02527-1>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02527-1>

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

- CAREER NEWS
- 20 September 2021

# Cash boost looms for historically Black US colleges and universities

Legal wins and federal budget proposals could address years of underfunding.

- [Chris Woolston](#) <sup>0</sup>



Hillary-Rhys Richard is part of an Apple-funded initiative at Huston-Tillotson University, a Texas HBCU, to train Black men as teachers.Credit: Apple

Historically Black colleges and universities (HBCUs) in the United States are set to ramp up investments in scientific research and education following a wave of new grants and proceeds from legal settlements.

Most of these institutions were created in the nineteenth century, to serve students of African descent. They are also hoping for extra funding as part of US President Joe Biden's proposed federal budget for fiscal year 2022.

There are more than 100 HBCUs in the United States. Around half are public institutions that rely mostly on government funding, but they have long been hampered by a lack of financial support, says Willie May, vice-president for research and economic development at Morgan State University, an HBCU in Baltimore, Maryland.

The extra funding should help HBCUs to increase their research output and eventually make the scientific workforce more diverse, he says. "This is a unique time. It looks like we're about to get a more equitable slice of the pie. We need to take advantage of it."

According to the US National Science Foundation (NSF), federal support for science and engineering activities at HBCUs [dropped by 37% between 2009 and 2019](#). By contrast, overall science and engineering funding for US higher-education institutions declined by just 10% over the same period.

But in May 2021, a federal judge approved a US\$577-million settlement for the four HBCUs in Maryland, including Morgan State. A lawsuit filed in 2006 had argued that the state had unfairly supported programmes and degrees at other public universities that were in competition with HBCUs.

The settlement will be divided between the institutions on the basis of student numbers. May estimates that his university will receive between \$150 million and \$200 million over 10 years, money that should help it to fulfil a key goal: progressing from its current Carnegie Classification as an R2 research institution (a doctoral university with high research activity) to an R1 (a doctoral university with very high research activity, the highest such classification) within the next decade. The move, he says, would boost the global profile of the university and help promote international outreach and collaboration.

In April, a legislative panel found that the state of Tennessee owed as much as \$544 million to Tennessee State University (TSU) in Nashville to remedy years of underfunding. Ebony McGee, a science-education researcher at Vanderbilt University in Nashville, says the state was essentially taking money from TSU and giving it to the University of Tennessee, Knoxville, a predominantly white institution.

## Federal funding

HBCUs across the country are also hoping for a new infusion of federal dollars. On 8 September, Democrats in the US House of Representatives proposed an educational budget for fiscal year 2022 over four years that includes \$1.45 billion for HBCUs, tribal colleges (those that are owned and operated by Indigenous communities on Indigenous property) and other institutions where most students and trainees are from minority ethnic groups.

Negotiations continue, but that level of funding would represent a substantial increase to the roughly \$1 billion the US government already invests in HBCUs each year, through mechanisms including tuition grants and research contracts.

The funding boost could have been much larger. In March, the Biden administration proposed to spend \$20 billion to upgrade laboratories and infrastructure for science, technology, engineering and mathematics (STEM) research at HBCUs. However, that provision was absent from the version of the bill that was passed by the US Senate in August.



Willie May hopes to make good use of increased funding at Morgan State University in Baltimore, Maryland. Credit: Morgan State University

In May, the US Department of Energy announced that it would spend \$17.3 million to create research opportunities and scholarships, with a focus on students of colour; much of the money will go directly to HBCUs. Howard University, a prominent HBCU in Washington DC, will receive nearly \$400,000 to support research on converting fossil fuels to hydrogen using electromagnetic energy.

Corporations are also showing new support for HBCUs. In June, Howard, Morgan State and two other historically Black institutions — Alabama A&M University in Normal and Prairie View A&M University in Texas — each received \$5-million “innovation grants” from Apple. Among other things, the grants are intended to support courses in computer architecture and silicon engineering.

Ten HBCUs have also received \$5 million each from Google as part of the company’s commitment to “address the diversity gaps in tech”, according to a company statement released in June.

## Expand and diversify

In its quest to become an R1 institution, May says, Morgan State intends to identify and invest in specific niches — including cybersecurity and coastal science — in which it could become a nationally recognized centre of excellence. “We’re happy with the settlement, but we have to be judicious,” he says.

As part of its commitment to STEM, the university has established new degree offerings in 2021, including a PhD in cybersecurity, a master’s in advanced computing and undergraduate degrees in cloud computing and mechanical engineering. It has also partnered with Purdue University in West Lafayette, Indiana, to offer undergraduate degrees in aeronautical and astronautical engineering.

According to the NSF, HBCUs confer [nearly one-fifth of all scientific bachelor's degrees](#) earned by Black students in the United States. And undergraduate education at an HBCU is often a path to an advanced degree: about one-third of Black PhD recipients in the sciences earned an undergraduate degree at an HBCU.



Ebony McGee says historically Black colleges and universities are good incubators of scientific talent. Credit: Ebony McGee

Eleven HBCUs are currently classified as R2 institutions, a number that has steadily grown over the years. “There was a trend toward more HBCUs becoming research institutions before this wave of funding started,” says Ivory Toldson, who studies counselling psychology at Howard and is director of education innovation and research at the US National Association for the Advancement of Colored People, a civil-rights organisation based in Baltimore, Maryland. No HBCU has yet reached R1 status, but Morgan State isn’t the only one looking to get there. Toldson says that administrators at Howard and at Texas Southern University in Houston hope eventually to move up to R1.

HBCUs can be great incubators for scientific talent, especially with sufficient resources, says McGee, who wrote a 2020 report on structural racism in STEM higher education ([\*E. O. McGee Educ. Res. 49, 633–644; 2020\*](#)). She notes that the demographic make-up of faculties at HBCUs tends to reflect that of the student body: “I don’t know a single HBCU that doesn’t have at least five black STEM faculty members.” That gives Black students a clearer vision of their possibilities than they often get at predominantly white institutions. “If you’ve never seen a Black engineering professor,” says McGee, “it’s hard to imagine yourself being a Black engineering professor.”

She adds that some Black students feel more supported and comfortable at an HBCU. “They want to be able to walk across campus without being stopped by the police,” she says.

HBCUs enrol many students who could have their pick of campuses, McGee says. But she also points out that they accept many who might have difficulty finding places elsewhere, because they are the first in their families to enter higher education, or they come from other under-privileged backgrounds.

Funds from settlements and one-time grants are welcome, but HBCUs will ultimately need more reliable support if they are going to flourish and compete, McGee says. She adds that HBCUs generally don’t have the endowments or wealthy donor bases that many predominantly white institutions enjoy. The American Council on Education reported in 2019 that endowments at public and private HBCUs are about 70% smaller, on

average, than those at other institutions. The report said that this disparity “jeopardizes an HBCU’s ability to buffer decreases in state and federal funding”.

“This whole funding extravaganza is two to three years old. These things often come in a wave and leave just as quickly,” says McGee. “We need the government and taxpayer dollars to support HBCUs robustly.”

*Nature* **597**, 578-579 (2021)

doi: <https://doi.org/10.1038/d41586-021-02528-0>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02528-0>

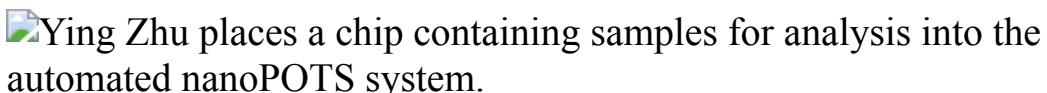
| [Section menu](#) | [Main menu](#) |

- TECHNOLOGY FEATURE
- 20 September 2021

# Single-cell proteomics takes centre stage

Deducing the full protein complement of individual cells has long played second fiddle to transcriptomics. That's about to change.

- [Jeffrey M. Perkel](#)



Chemist Ying Zhu places a nanoPOTS chip containing protein samples into an automated analysis system. Credit: Andrea Starr/PNNL

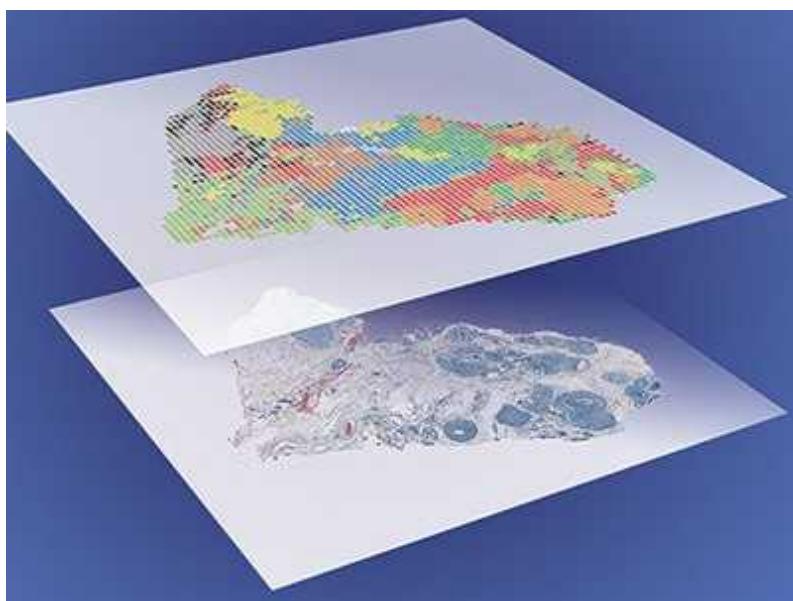
Claudia Cortecka was both sceptical and intrigued when her thesis adviser told her in 2018 about a new method that uses mass spectrometry to analyse the protein contents of individual cells. When he said he was looking for someone to pursue this single-cell proteomics strategy in his laboratory at the Vienna BioCenter research institute, she decided to take a chance.

“I was always very much interested in mass spectrometry,” she says. “And I thought, ‘why not go for the challenge?’ I wanted to look into that [strategy] a bit deeper and closer.”

With no backup plan, the project was a sink-or-swim proposition, Cortecka says. “It was basically just: do single cell, make it work, or try harder.” Yet work it did. In April, she and her colleagues detailed a new sample-preparation device called proteoCHIP, which they used to map some 2,000 proteins across 158 single cells from 2 human cell types<sup>1</sup>.

That study is one of at least half a dozen over the past year that have described single-cell proteomics strategies, tools and preliminary findings. And more are coming. In 2018, Nikolai Slavov, a systems biologist at Northeastern University in Boston, Massachusetts, hosted his first annual conference on single-cell proteomics, which attracted about 50 attendees. This year's (mostly virtual) conference had more than 1,300. "The growth has been exponential," he says.

Most single-cell studies focus on nucleic acids, especially the transcriptome — which represents all the expressed genes in a cell. But proteins, says Neil Kelleher, a biochemist at Northwestern University in Evanston, Illinois, are "the worker bees" of the cell. "The amounts, the post-translational modifications, the proteoform dynamics — this is closer to the phenotype," he explains. "And that means that disease diagnostics, response to drugs, all the human biology we want to engage with — to control, steer, detect — it needs proteomics."



### Single-cell analysis enters the multiomics age

Proteomics aims to catalogue and characterize the total complement of protein isoforms from a cell, tissue, organ or organism. (These 'proteoforms' are encoded by the same gene but have non-identical amino-acid sequences or post-translational modifications.) However, at the single-cell level, that's easier said than done. Each type of nucleic acid behaves largely in a

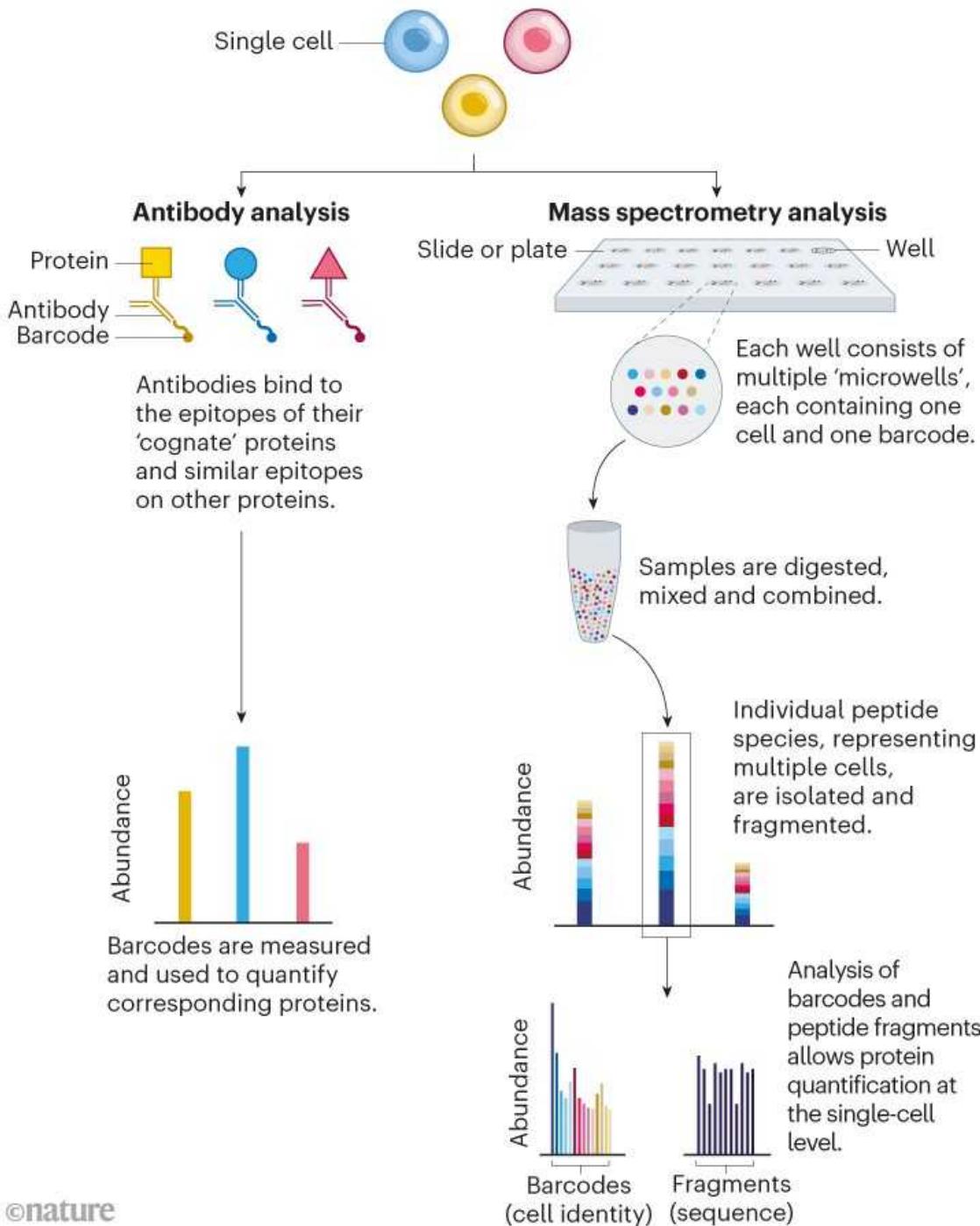
predictable way. But the proteome has a vast array of different chemistries, interactions, dynamics and abundances. And with no protein equivalent to PCR amplification of DNA, any technique to detect proteins must be sensitive enough to identify them, however little material a cell contains.

Using antibodies, that's relatively straightforward. Flow cytometry and mass cytometry, for instance, can each quantify up to about 50 proteins per cell. And high-resolution microscopy, as used in the [Human Protein Atlas project](#), intrinsically provides single-cell resolution.

But not all proteins have corresponding antibodies, and some antibodies bind to proteins only weakly or non-specifically. Furthermore, because antibody-based approaches target specific proteins, researchers can only see that portion of the proteome. Many in the single-cell proteomics community have instead turned to mass spectrometry, a non-targeted method that identifies and quantifies molecules on the basis of their mass and charge (see 'Two paths to the proteome').

## TWO PATHS TO THE PROTEOME

Researchers can probe single-cell protein content using antibodies or mass spectrometry. The antibody-based approach is limited in the number of proteins it can handle at once, and mass spectrometry requires efficiently collecting and quantifying the contents of single cells.



Source: Adapted from N. Slavov *Science* **367**, 512–513 (2020).

The fact that mass spectrometry is sensitive enough to identify at least some proteins at the single-cell level was never in doubt: some instruments can detect attomolar ( $10^{-18}$  moles) quantities of material, the equivalent of several hundred thousand ions. According to one study<sup>2</sup>, the median mammalian protein is present at 18,000 copies per cell. But manipulating the contents of a single cell and faithfully transferring them into a mass spectrometer poses challenges.

As recently as five years ago, says Matthias Mann, director of the Max Planck Institute of Biochemistry in Munich, Germany, “the community was so far away from single-cell sensitivity, and also from handling single cells” that he used to think “it might happen some time, but not in my career”. Yet the field has accelerated faster than he expected.

## Small samples

According to Mann, that acceleration stems not only from advances in instrumentation and analytical tools, but also, crucially, in sample preparation. “You want to have this whole reaction happen in a small volume so that you don’t lose the proteins and they don’t adsorb everywhere,” he explains.

Ctortecka’s proteoCHIP is one such design; another is nanoPOTS, developed by chemists Ryan Kelly and Ying Zhu at the Pacific Northwest National Laboratory in Richland, Washington.

NanoPOTS is like a nanolitre-scale microtiter plate fabricated onto the surface of a microscope slide<sup>3</sup>. Each ‘well’ is a hydrophobic circle about one millimetre in diameter, with a small hydrophilic ‘pedestal’ at the centre at which cells are deposited and prepared. “Think about the mesas in Arizona,” says Kelly, now at Brigham Young University in Provo, Utah, referring to the US state’s iconic, flat-topped hills: “All the stuff is taking place on the top of the mesa.”

The pedestal’s small area means there is a smaller surface for proteins to adhere to — about 99.5% less than a 0.5-millilitre centrifuge tube, as Kelly and Zhu note in their study. The correspondingly small reaction volumes

(less than 200 nanolitres) increase enzyme concentration and thus efficiency. Add the fact that the reaction protocol limits liquid-handling steps, and the result is an increased yield of proteins per cell. Kelly's team observed from 2- to 25-fold more peptides with nanoPOTS than when samples were prepared in 0.5-millilitre centrifuge tubes. Using nanoPOTS, Kelly's team has detected an average of 1,085 and 1,012 proteins for each of two classes of primary human neuron<sup>4</sup>.



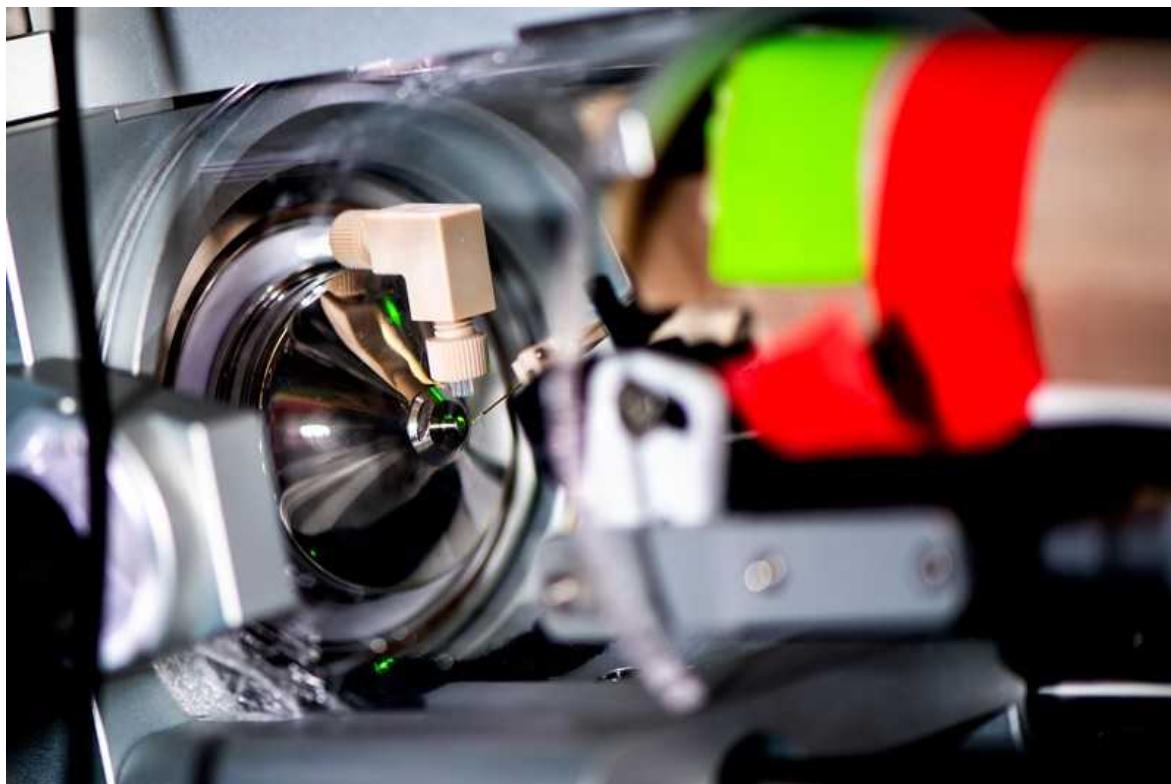
### A dream of single-cell proteomics

How comprehensive that is depends on how you count — some genes encode multiple proteoforms, for instance, and not all proteins are expressed in all cells. However, that number is par for the course for single-cell proteomics: some researchers claim to have improved on it in unpublished work, but most studies identify about 1,000 proteins per cell (although the total number of identified proteins across all cells is higher). In a February preprint<sup>5</sup>, for instance, Mann's team used a new instrument design from mass-spectrometry vendor Bruker in Billerica, Massachusetts, to detect proteome differences as cells progress through the cell cycle. The median number of proteins detected per cell-cycle stage ranged from 611 in the growth phase of cell division to 1,263 in the subsequent phase, when DNA is synthesized. Ongoing work has detected more (952 and 1,773, respectively). But that number was enough to tease apart biological differences. “Every single cell has quite a stable proteome,” Mann notes,

meaning that researchers might be able to analyse fewer cells than other single-cell methods require. “Conceptually, that is the most exciting result of that paper,” Mann says.

It still takes a long time to acquire those data, however. Single-cell proteomics studies tend to use ‘bottom-up’ strategies to identify proteins from a smattering of peptide fragments rather than looking for intact proteins. But those peptides are identified one at a time, not in parallel. And the mass spectrometer needs time to accumulate each ion. For one study, Erwin Schoof, a biological mass spectrometrist at the Technical University of Denmark in Lyngby, allocated half a second per peptide in a 160-minute run. “On a good day we are measuring 4,500 peptides per cell,” Schoof says. As a result, his team could analyse just eight samples per day.

Sample preparation is also a bottleneck. With 27 wells, the original nanoPOTS could process 27 single cells at a time. Zhu’s second-generation ‘nested nanoPOTS’ (N2) design contains a  $3 \times 3$  grid of pedestals in each well, supporting up to 243 cells ( $27 \times 9$ ) at once<sup>6</sup>. According to Zhu, N2 was designed to accommodate another crucial development in single-cell proteomics: multiplexing, which increases throughput. 

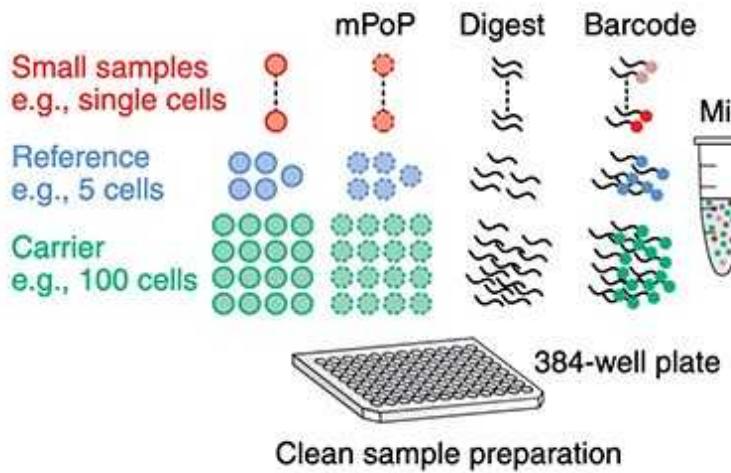


Part of the mass spectrometer used by systems biologist Nikolai Slavov to study single-cell proteomics. Credit: Northeastern University/Ruby Wallau

In 2018, Slavov's team described a method called SCoPE-MS (single-cell proteomics by mass spectrometry)<sup>7</sup>, which blends a mass-spectrometer-friendly cell-lysis protocol with a protein carrier that increases the amount of material available for sequencing. "This kind of approach immediately increased our ability to determine peptide sequences without doing anything difficult," Slavov explains. "We were outsmarting the problem rather than brute-forcing it."

## Barcode breakthrough

Crucially, SCoPE-MS also features mass spectrometry's version of barcoding: isobaric tags. These are molecules with identical masses that fragment into differently sized ions inside a mass spectrometer. By coupling different tags to different samples, researchers can work out how much of a given protein is present in each one. Using tandem mass tag (TMT) reagents, for instance, researchers can differentiate between up to 18 samples in a single mixture<sup>8</sup>. But to do so, the samples must be labelled individually and then pooled — a technically challenging step, given the small volumes involved. "The robot has to be very precise to withdraw this nanolitre volume and put them together for mass-spectrometer analysis," Zhu explains. N2 allows researchers to process cells individually but then pool them in a single step by adding a large enough droplet of buffer to cover all the individual pedestals in one 'well', thus circumventing that issue.



### Towards resolving proteomes in single cells

At this year's Single Cell Proteomics conference in Boston, Slavov's graduate student Andrew Leduc presented an alternative approach. Leduc described how he and his colleagues used a piezo-acoustic dispensing device to array and prepare some 1,500 cells in 20-nanolitre droplets. These were clustered in groups of 12–14 on microscope slides to simplify pooling, and surrounded by a perimeter of water droplets to increase humidity and prevent evaporation<sup>9</sup>. The team has used that method to study macrophage stimulation and the cell division cycle.

Meanwhile, other members of Slavov's team have revamped SCoPE-MS. SCoPE2 uses a simpler cell-lysis approach and improved analysis pipeline<sup>10</sup>, and is broadly accessible and scalable for production use, Slavov says.

Other researchers are trying to make the most of their instruments' precious time by changing how they collect data. Most mass spectrometrists run their machines in a 'data-dependent acquisition' mode, in which the instrument identifies and sequences the most abundant ions. As a result, these analyses tend to overlook the most interesting, lower-abundance proteins.

Another option is a targeted approach, in which the instrument is told specifically which ions to look for. But some researchers are now exploring strategies that scan everything in the sample and work out the details later.

These ‘data-independent analysis’ methods are not typically compatible with multiplexing, but in February, Ctortecka and her colleagues reported a strategy for combining the two<sup>11</sup>. “So you have a systematic way to look at your peptides in your sample, and this is performed in every single run exactly the same,” she says.

For his part, Schoof says he is working with vendors to accelerate chromatographic separations, and thus speed up experiments from 160 minutes to an hour. Using other optimizations, he has a roadmap to ramp up to 20 samples, or 360 multiplexed cells, per day. At that rate, he says, “a 10,000-cell experiment like you see in single-cell RNA-seq is, for lack of a better word, ‘only’ one month of runtime. In terms of doing single-cell proteomics, that’s already quite an achievement.”

## Another dimension

Most single-cell methods remove cells from their tissue context. But where in the tissue a cell resides actually matters. By disaggregating cells, researchers lose what Mann calls their “sociology”. So he and other researchers are working to add a spatial dimension to single-cell proteomics, although none of the approaches is yet at the single-cell level.

Last year, Kelly and his colleagues published a strategy combining nanoPOTS, laser-capture microdissection (which uses a laser to excise cells from tissue) and mass spectrometry to detail some 2,000 proteins per 100-micrometre pixel<sup>12</sup>. In May, a team led by cancer researcher Thomas Cox of the Garvan Institute of Medical Research in Sydney, Australia, and vision scientist Gus Grey at the University of Auckland, New Zealand, combined ultra-high-resolution mass spectrometry and an R software package called HIT-MAP to sequence and identify proteins in intact samples of bovine lens tissue<sup>13</sup>.



## NatureTech hub

And in January, Mann and his team reported a strategy called Deep Visual Proteomics<sup>14</sup>, which blends artificial intelligence, microscopy and laser-capture microdissection to automatically identify, isolate and characterize as few as 100 cells of a given type in tissue. His team used the approach to differentiate between cells at the centre and periphery of human melanoma samples. “I think this can be quite a game-changer,” he says.

Others, such as Kelleher, are pushing for single-molecule, single-cell proteomics — that is, the ability to sequence individual protein molecules in a cell. At the moment, he says, “we’re barely at proof-of-concept for some of these underlying technologies.” But their development is likely to get a boost. In July, the US National Institutes of Health announced some US\$20 million in funding for technology development in single-molecule and single-cell proteomics. And Kelleher estimates that private investors have poured some \$2 billion more into the subfield.

To make the most of those technologies, Kelleher and others advocate for a comprehensive atlas of all the human proteoforms that could be present in a sample. Just as the Human Genome Project provided a reference genome that made next-generation DNA sequencing technologies more powerful, Kelleher and his colleagues envision a Human Proteoform Project to create what they call “a definitive reference set of the proteoforms produced from

the genome”<sup>15</sup>. Such a resource could enhance the power of both single-cell and single-molecule proteomics technologies by allowing researchers to concentrate more on ‘scoring’ proteins than discovering them, Kelleher says.

There’s no guarantee that such an atlas will come to pass. But when it comes to ’omics, one should never bet against the technology. When she started her doctoral work, Ctortecka doubted her project would succeed, but thought she would learn something interesting in any event. “I was very much convinced that this would never be possible,” she says. “Look where we are now.”

*Nature* **597**, 580-582 (2021)

doi: <https://doi.org/10.1038/d41586-021-02530-6>

## References

1. 1.

Hartlmayr, D. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.04.14.439828> (2021).

2. 2.

Nagaraj, N. *et al.* *Mol. Syst. Biol.* **7**, 548 (2011).

3. 3.

Zhu, Y. *et al.* *Nature Commun.* **9**, 882 (2018).

4. 4.

Cong, Y. *et al.* *Chem. Sci.* **12**, 1001–1006 (2021).

5. 5.

Brunner, A.-D. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2020.12.22.423933> (2021).

6. 6.

Woo, J. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.02.17.431689> (2021).

7. 7.

Budnik, B., Levy, E., Harmange, G. & Slavov, N. *Genome Biol.* **19**, 161 (2018).

8. 8.

Li, J. *et al.* *J. Proteome Res.* **20**, 2964–2972 (2021).

9. 9.

Leduc, A., Huffman, R. G. & Slavov, N. Preprint at bioRxiv  
<https://doi.org/10.1101/2021.04.24.441211> (2021).

10. 10.

Specht, H. *et al.* *Genome Biol.* **22**, 50 (2021).

11. 11.

Ctortecka, C. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.02.11.430601> (2021).

12. 12.

Piehowski, P. D. *et al.* *Nature Commun.* **11**, 8 (2020).

13. 13.

Guo, G. *et al.* *Nature Commun.* **12**, 3241 (2021).

14. 14.

Mund, A. *et al.* Preprint at bioRxiv  
<https://doi.org/10.1101/2021.01.25.427969> (2021).

15. 15.

Smith, L. M. *et al.* Preprint at  
<https://doi.org/10.20944/preprints202010.0368.v1> (2020).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02530-6>

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

- WHERE I WORK
- 20 September 2021

# Preserving pieces of history in eggshells and birds' nests

Museum curator Douglas Russell catalogues and maintains specimens that offer a glimpse into the breeding behaviours of birds both living and extinct.

- [Richa Malhotra](#) <sup>0</sup>



Douglas Russell is senior curator at the Natural History Museum at Tring, UK. Credit: Alecsandra Dragoi for *Nature*

Here at the Natural History Museum at Tring, UK, I'm in our nest collection, which numbers just over 4,000. Behind me are 67 metal cabinets with nests arranged in taxonomic order. Each nest is labelled with the date and place of collection, and the collector's name. Next to me is a 1928 mud nest from Argentina that was made by the rufous hornero (*Furnarius rufus*), known for its large, globular nests that shield eggs and young from predators.

I'm the senior curator of birds' eggs and nests. I ensure that specimens are stored appropriately to prevent damage and are well catalogued, so we know exactly what we have and where. Our nest and egg collections are the most comprehensive archive of information on bird breeding in the world. When I came here about 20 years ago, the nest collection was rarely used and we didn't know how many examples of extinct and endangered species we had. I've spent a lot of time and effort cataloguing and understanding

these particular 129 nests, 40 of which belong to extinct birds such as the Laysan crake (*Zapornia palmeri*) and the Aldabra brush warbler (*Nesillas aldabrana*).

We have up to 300,000 sets of eggs. I am holding four dunlin (*Calidris alpina*) eggshells, collected in 1952 in Ireland. They were donated to the Wildfowl & Wetlands Trust, a UK conservation charity, which gave them to us as part of a larger collection.

I have been interested in birds and natural history since childhood, and my mother used to take me to the Royal Museum of Scotland (now the National Museum of Scotland) in Edinburgh. After graduating in biological sciences from Edinburgh Napier University, I volunteered at the museum before getting my first paid museum job.

When researchers want to access the collections, I check that we have specimens relevant to their research, discuss exactly what they intend to do and work with them to minimize the risk of damage. Although I want our collections to result in robust science, they must be preserved.

*Nature* **597**, 586 (2021)

doi: <https://doi.org/10.1038/d41586-021-02529-z>

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02529-z>

# Research

- **Modern Polynesian genomes offer clues to early eastward migrations** [ 22 September 2021]

News & Views • A genome-wide analysis of modern populations in Polynesia suggests the direction and timing of ancient Polynesian migrations. This model bears consistencies and inconsistencies with models based on archaeology and linguistics.

- **Ebola virus can lie low and reactivate after years in human survivors** [ 15 September 2021]

News & Views • A genomic comparison of Ebola virus from the 2021 outbreak in Guinea with sequences from the West African outbreak that ended in 2016 suggests that the virus can remain latent in human survivors for an extended period of time.

- **Seed-inspired vehicles take flight** [ 22 September 2021]

News & Views • Many plant seeds have shapes that aid their efficient dispersal by wind. Inspired by these seeds, a range of fliers have been constructed that could have applications from environmental monitoring to wireless communication.

- **Policy, drought and fires combine to affect biodiversity in the Amazon basin** [ 01 September 2021]

News & Views • Analysis of the ranges of nearly 15,000 plant and vertebrate species in the Amazon basin reveals that, from 2001 to 2019, a majority were affected by fire. Drought and forest policy were the best predictors of fire outcomes.

- **Light detection nears its quantum limit** [ 22 September 2021]

News & Views • Organic molecules are increasingly crucial in quantum-optics technologies. An experiment shows how the strong coupling between confined organic molecules and light can improve photon detection at room temperature.

- **Quenching of star formation from a lack of inflowing gas to galaxies** [ 22 September 2021]

Article • The authors report 1.3 mm observations of dust emission from strongly lensed galaxies where star formation is quenched, demonstrating that gas depletion is responsible for the cessation of star formation in some high-redshift galaxies.

- **Normal, dust-obscured galaxies in the epoch of reionization**

[ 22 September 2021]

Article • Two serendipitously detected dust-obscured galaxies are reported at  $z = 6.7$  and  $7.4$ , with estimates that such galaxies provide an additional 10–25% contribution to the total star formation rate density at  $z > 6$ .

- **Single-photon nonlinearity at room temperature** [ 22 September 2021]  
Article • Nonlinearity induced by a single photon is desirable because it can drive power consumption of optical devices to their fundamental quantum limit, and is demonstrated here at room temperature.
- **Electron phase-space control in photonic chip-based particle acceleration** [ 22 September 2021]  
Article • In a tiny chip-based particle accelerator, phase-space control of the emerging electron beam demonstrates guiding over a length of nearly 80 micrometres and an indispensable prerequisite to electron acceleration to high energies.
- **Three-dimensional electronic microfliers inspired by wind-dispersed seeds** [ 22 September 2021]  
Article • With a design inspired by wind-dispersed seeds, a series of three-dimensional passive fliers at the macro-, meso- and microscale are realized that can bear active electronic payloads.
- **Three-dimensional magnetic stripes require slow cooling in fast-spread lower ocean crust** [ 22 September 2021]  
Article • A record of Earth's magnetic field constructed from near-bottom magnetization observations and oriented samples provides three-dimensional imaging of magnetic stripes in fast-spread crust, and suggests slow cooling off-axis, as opposed to deep hydrothermal cooling close to the spreading ridge.
- **How deregulation, drought and increasing fire impact Amazonian biodiversity** [ 01 September 2021]  
Article • Remote-sensing estimates of fires and the estimated geographic ranges of thousands of plant and vertebrate species in the Amazon Basin reveal that fires have impacted the ranges of 77–85% of threatened species over the past two decades.
- **Paths and timings of the peopling of Polynesia inferred from genomic networks** [ 22 September 2021]  
Article • Analysis of genomic networks from 430 modern individuals across 21 Pacific island populations reveals the human settlement history of Polynesia.
- **Rare variant contribution to human disease in 281,104 UK Biobank exomes** [ 10 August 2021]  
Article • The authors analyse rare protein-coding genetic variants for association with 18,780 traits in the UK Biobank cohort.
- **Bioaccumulation of therapeutic drugs by human gut bacteria** [ 08 September 2021]  
Article • An analysis of the interactions between 15 drugs and 25 gut bacterial strains shows that bioaccumulation of drugs within bacterial cells is another mechanism through which gut

microorganisms can alter drug availability and efficacy.

- **Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks** [ 15 September 2021]

Article • The viral lineage responsible for the February 2021 outbreak of Ebola virus disease in Guinea is nested within a clade that predominantly consists of genomes sampled during the 2013–2016 epidemic, suggesting that the virus might have re-emerged after a long period of latency within a previously infected individual.

- **An engineered IL-2 partial agonist promotes CD8+ T cell stemness** [ 15 September 2021]

Article • H9T, an engineered IL-2 partial agonist, promotes the expansion of T cells while maintaining a stem-cell-like state, leading to improved efficacy of adoptive cell therapy in mouse models of melanoma and acute lymphoblastic leukaemia.

- **Inter-cellular CRISPR screens reveal regulators of cancer cell phagocytosis** [ 08 September 2021]

Article • CRISPR-mediated gene knockout and activation are used to identify molecules in cancer cells and macrophages that regulate antibody-dependent cancer cell phagocytosis.

- **Using DNA sequencing data to quantify T cell fraction and therapy response** [ 08 September 2021]

Article • A robust, cost-effective technique based on whole-exome sequencing data can be used to characterize immune infiltrates, relate the extent of these infiltrates to somatic changes in tumours, and enables prediction of tumour responses to immune checkpoint inhibition therapy.

- **Single-cell Ribo-seq reveals cell cycle-dependent translational pausing** [ 08 September 2021]

Article • Highly sensitive ribosome profiling of single cells at single-codon resolution enables identification of distinct cell cycle-dependent translational dynamic states in individual cells.

- **Structural basis for tRNA methylthiolation by the radical SAM enzyme MiaB** [ 15 September 2021]

Article • Crystal structures reveal the catalytic mechanism through which the radical S-adenosylmethionine enzyme MiaB adds a methylthio group onto tRNA.

- **Positive allosteric mechanisms of adenosine A1 receptor-mediated analgesia** [ 08 September 2021]

Article • MIPS521, a positive allosteric modulator of the adenosine A1 receptor, has analgesic properties in a rat model of neuropathic pain through a mechanism by which MIPS521 stabilizes the complex between adenosine, receptor and G protein.

- **A finding of sex similarities rather than differences in COVID-19 outcomes** [ 22 September 2021]

Matters Arising •

- **Reply to: A finding of sex similarities rather than differences in COVID-19 outcomes** [ 22 September 2021]

Matters Arising •

- NEWS AND VIEWS
- 22 September 2021

# Modern Polynesian genomes offer clues to early eastward migrations

A genome-wide analysis of modern populations in Polynesia suggests the direction and timing of ancient Polynesian migrations. This model bears consistencies and inconsistencies with models based on archaeology and linguistics.

- [Patrick V. Kirch](#)    [ORCID: http://orcid.org/0000-0003-4264-6689](#) <sup>0</sup>

Determining the timing and sequence of people's discovery of and settlement in their island homes in the Pacific Ocean has been an enduring scientific problem. Genetic studies published in the past few years have determined the history of human movements across the southwestern Pacific<sup>1</sup>, but the settlement of Polynesia remains the subject of ongoing debate. In a [methodologically innovative study](#), Ioannidis *et al.*<sup>2</sup> use genomic data to propose a sequence of migrations, starting in Samoa and progressing rapidly eastwards through the Southern Cook Islands in the ninth century, thence to the Society Islands and Tuamotu Islands, and finally, by the mid-fourteenth century, to the widely separated islands of the Marquesas, Raivavae and Rapa Nui (also known as Easter Island).

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

## Rent or Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **597**, 477–478 (2021)

doi: <https://doi.org/10.1038/d41586-021-01719-z>

## **References**

1. 1.

Springer, M. & Reich, D. *World Archaeol.* **51**, 620–639 (2020).

2. 2.

Ioannidis, A. G. *et al. Nature* <https://doi.org/10.1038/s41586-021-03902-8> (2021).

3. 3.

Kirch, P. V. *Annu. Rev. Anthropol.* **39**, 131–148 (2010).

4. 4.

Niespolo, E. M., Sharp, W. D. & Kirch, P. V. *J. Archaeol. Sci.* **101**, 21–33 (2019).

5. 5.

Kirch, P. V., Molle, G., Niespolo, E. M. & Sharp, W. D. *J. Archaeol. Sci. Rep.* **35**, 102724 (2021).

6. 6.

Kirch, P. V. *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact* 2nd edn, 201 (Univ. California Press, 2017).

7. 7.

Conte, E. & Molle, G. *Archaeol. Oceania* **49**, 121–136 (2014).

8. 8.

Walworth, M. *Ocean. Linguist.* **53**, 256–272 (2014).

9. 9.

Weisler, M. I. *et al. Proc. Natl Acad. Sci. USA* **113**, 8150–8155 (2016).

10. 10.

Kirch, P. V., Carpenter, A. & Ruggles, C. *Rapa Nui J.* **32**, 37–57 (2019).

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

- NEWS AND VIEWS
- 15 September 2021

# Ebola virus can lie low and reactivate after years in human survivors

A genomic comparison of Ebola virus from the 2021 outbreak in Guinea with sequences from the West African outbreak that ended in 2016 suggests that the virus can remain latent in human survivors for an extended period of time.

- [Robert F. Garry](#) <sup>0</sup>

Infection by Ebola virus can be fatal, so understanding what drives human outbreaks might offer better ways to control it. [Writing in Nature](#), Keita *et al.*<sup>1</sup> provide evidence that the 2021 outbreak of Ebola virus in Guinea, West Africa, was triggered by reactivation of an infection that had been dormant in a person without evoking symptoms. Although reawakening of such clinically latent Ebola virus infections has been observed previously, the length of the latency period — nearly five years from the end of the 2013–16 West African Ebola outbreak — is surprising. The unexpected observation that the virus can persist in the human body for such a long time has considerable implications for public health and care of survivors of Ebola.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 478-480 (2021)

doi: <https://doi.org/10.1038/d41586-021-02378-w>

## **References**

1. 1.

Keita, A. K. *et al.* *Nature* **597**, 539–543 (2021).

2. 2.

Diehl, W. E. *et al.* *Cell* **167**, 1088–1098 (2016).

3. 3.

Kinganda-Lusamaki, E. *et al.* *Nature Med.* **27**, 710–716 (2021).

4. 4.

Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. *Adv. Virus Res.* **100**, 163–188 (2018).

5. 5.

Leroy, E. M. *et al.* *Nature* **438**, 575–576 (2005).

6. 6.

MacLean, O. A. *et al.* *PLoS Biol.* **19**, e3001115 (2021).

7. 7.

Lee, H. & Nishiura, H. *Int. J. Infect. Dis.* **64**, 90–92 (2017).

8. 8.

Lieberman, P. M. *Cell Host Microbe* **19**, 619–628 (2016).

9. 9.

Jacobs, M. *et al.* *Lancet* **388**, 498–503 (2016).

10. 10.

Varkey, J. B. *et al.* *N. Engl. J. Med.* **372**, 2423–2427 (2015).

11. 11.

Mbala-Kingebeni, P. *et al.* *N. Engl. J. Med.* **384**, 1240–1247 (2021).

12. 12.

Bond, N. G. *et al.* *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab267> (2021).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02378-w>

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

- NEWS AND VIEWS
- 22 September 2021

# Seed-inspired vehicles take flight

Many plant seeds have shapes that aid their efficient dispersal by wind. Inspired by these seeds, a range of fliers have been constructed that could have applications from environmental monitoring to wireless communication.

- [E. Farrell Helbling](#) 0

In [a paper in \*Nature\*](#), Kim *et al.*<sup>1</sup> report 3D fliers that are inspired by the passive, helicopter-style wind-dispersal mechanism of certain seeds. The adopted production processes enable the rapid parallel fabrication of many fliers and permit the integration of simple electronic circuits using standard silicon-on-insulator techniques. Tuning the design parameters — such as the diameter, porosity and wing type — generates beneficial interactions between the devices and the surrounding air. Such interactions lower the terminal velocity of the fliers, increase air resistance and improve stability by inducing rotational motion. When combined with complex integrated circuits, these devices could form dynamic sensor networks for environmental monitoring, wireless communication nodes or various other technologies based on the network of Internet-connected devices called the Internet of Things.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

## Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

## Rent or Buy

All prices are NET prices.

## **Additional access options:**

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

*Nature* **597**, 480-481 (2021)

doi: <https://doi.org/10.1038/d41586-021-02490-x>

## **References**

1. 1.

Kim, B. H. *et al.* *Nature* **597**, 503–510 (2021).

2. 2.

Lindsey, Q., Mellinger, D. & Kumar, V. *Auton. Robots* **33**, 323–336 (2012).

3. 3.

McGuire, K. N., De Wagter, C., Tuyls, K., Kappen, H. J. & de Croon, G. C. H. E. *Sci. Robot.* **4**, aaw9710 (2019).

4. 4.

Wood, R., Nagpal, R. & Wei, G.-Y. *Sci. Am.* **308**, 60–65 (2013).

5. 5.

Jafferis, N. T., Helbling, E. F., Karpelson, M. & Wood, R. J. *Nature* **570**, 491–495 (2019).

6. 6.

Karásek, M., Muijres, F. T., De Wagter, C., Remes, B. D. W. & de Croon, G. C. H. E. *Science* **361**, 1089–1094 (2018).

7. 7.

Ramezani, A., Chung, S.-J. & Hutchinson, S. *Sci. Robot.* **2**, aal2505 (2017).

8. 8.

Warneke, B., Last, M., Liebowitz, B. & Pister, K. S. J. *Computer* **34**, 44–51 (2001).

9. 9.

Piech, D. K. *et al.* *Nature Biomed. Eng.* **4**, 207–222 (2020).

10. 10.

Cortese, A. J. *et al.* *Proc. Natl Acad. Sci. USA* **117**, 9173–9179 (2020).

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

- NEWS AND VIEWS
- 01 September 2021

# Policy, drought and fires combine to affect biodiversity in the Amazon basin

Analysis of the ranges of nearly 15,000 plant and vertebrate species in the Amazon basin reveals that, from 2001 to 2019, a majority were affected by fire. Drought and forest policy were the best predictors of fire outcomes.

- [Thomas W. Gillespie](#) <sup>0</sup>

The Amazon basin contains the largest continuous area of tropical rainforests in the world, and has a crucial role in regulating Earth's climate<sup>1</sup>. Rates of tropical-rainforest deforestation and the impacts of fire and drought there are well established<sup>2,3</sup>. Less is known, however, about how these factors might interact to affect biodiversity, and about the role that forest policy and its enforcement have had over time. [Writing in Nature](#), Feng *et al.*<sup>4</sup> address these issues.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 481–483 (2021)

*doi:* <https://doi.org/10.1038/d41586-021-02320-0>

## **References**

1. 1.

Marengo, J. A., Tomasella, J., Soares, W. R., Alves, L. M. & Nobre, C. A. *Theor. Appl. Climatol.* **107**, 73–85 (2012).

2. 2.

Nepstad, D. C. *et al.* *Nature* **398**, 505–508 (1999).

3. 3.

Davidson, E. A. *et al.* *Nature* **481**, 321–328 (2012).

4. 4.

Feng, X. *et al. Nature* **597**, 516–521 (2021).

5. 5.

Nepstad, D. *Science* **344**, 1118–1123 (2014).

6. 6.

Hansen, M. C. *et al. Science* **342**, 850–853 (2013).

7. 7.

Libonati, R. *et al. Sci. Rep.* **11**, 4400 (2021).

8. 8.

Hopkins, M. J. G. *J. Biogeogr.* **34**, 1400–1411 (2007).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02320-0>

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

- NEWS AND VIEWS
- 22 September 2021

# Light detection nears its quantum limit

Organic molecules are increasingly crucial in quantum-optics technologies. An experiment shows how the strong coupling between confined organic molecules and light can improve photon detection at room temperature.

- [Sebastian Klembt](#) 0

Experiments in quantum optics aim to explain the intrinsic quantum properties of light. The past few decades have seen substantial improvements in both theory and experiment to understand, control and manipulate quantum states of light. Innovative nanotechnological techniques could enable a new generation of all-optical devices, such as switches and amplifiers, that operate at the fundamental quantum limit. At this limit, quantum optics have helped to launch the field of quantum technologies, in which quantum states of light lie at the core of transformative technological applications. [Writing in Nature](#), Zasedatelev *et al.*<sup>1</sup> report an innovative way to use phenomena called optical nonlinearities in organic microcavities (light-trapping structures) that allows light detection at the single-photon level in ambient conditions.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

### Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

### Rent or Buy

All prices are NET prices.

### **Additional access options:**

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

*Nature* **597**, 483–484 (2021)

doi: <https://doi.org/10.1038/d41586-021-02489-4>

## References

1. 1.

Zasedatelev, A. V. *et al.* *Nature* **597**, 493–497 (2021).

2. 2.

Chikkaraddy, R. *et al.* *Nature* **535**, 127–130 (2016).

3. 3.

Maser, A., Gmeiner, B., Utikal, T., Götzinger, S. & Sandoghdar, V. *Nature Photon.* **10**, 450–453 (2016).

4. 4.

Carusotto, I. & Ciuti, C. *Rev. Mod. Phys.* **85**, 299–366 (2013).

5. 5.

Kéna-Cohen, S. & Forrest, S. R. *Nature Photon.* **4**, 371–375 (2010).

6. 6.

Plumhof, J. D., Stöferle, T., Mai, L., Scherf, U. & Mahrt, R. F. *Nature Mater.* **13**, 247–252 (2014).

7. 7.

Zasedatelev, A. V. *et al.* *Nature Photon.* **13**, 378–383 (2019).

8. 8.

He, Y.-M. *et al.* *Nature Nanotech.* **8**, 213–217 (2013).

9. 9.

Somaschi, N. *et al.* *Nature Photon.* **10**, 340–345 (2016).

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02489-4>

- Article
- [Published: 22 September 2021](#)

# Quenching of star formation from a lack of inflowing gas to galaxies

- [Katherine E. Whitaker](#) ORCID: [orcid.org/0000-0001-7160-3632<sup>1,2</sup>](https://orcid.org/0000-0001-7160-3632),
- [Christina C. Williams](#) ORCID: [orcid.org/0000-0003-2919-7495<sup>3</sup>](https://orcid.org/0000-0003-2919-7495),
- [Lamiya Mowla](#) ORCID: [orcid.org/0000-0002-8530-9765<sup>4</sup>](https://orcid.org/0000-0002-8530-9765),
- [Justin S. Spilker<sup>5</sup>](#),
- [Sune Toft](#) ORCID: [orcid.org/0000-0003-3631-7176<sup>2,6</sup>](https://orcid.org/0000-0003-3631-7176),
- [Desika Narayanan<sup>2,7</sup>](#),
- [Alexandra Pope](#) ORCID: [orcid.org/0000-0001-8592-2706<sup>1</sup>](https://orcid.org/0000-0001-8592-2706),
- [Georgios E. Magdis](#) ORCID: [orcid.org/0000-0002-4872-2294<sup>2,6,8</sup>](https://orcid.org/0000-0002-4872-2294),
- [Pieter G. van Dokkum<sup>9</sup>](#),
- [Mohammad Akhshik<sup>10</sup>](#),
- [Rachel Bezanson](#) ORCID: [orcid.org/0000-0001-5063-8254<sup>11</sup>](https://orcid.org/0000-0001-5063-8254),
- [Gabriel B. Brammer](#) ORCID: [orcid.org/0000-0003-2680-005X<sup>2,6</sup>](https://orcid.org/0000-0003-2680-005X),
- [Joel Leja](#) ORCID: [orcid.org/0000-0001-6755-1315<sup>12,13,14</sup>](https://orcid.org/0000-0001-6755-1315),
- [Allison Man](#) ORCID: [orcid.org/0000-0003-2475-124X<sup>15</sup>](https://orcid.org/0000-0003-2475-124X),
- [Erica J. Nelson<sup>16</sup>](#),
- [Johan Richard<sup>17</sup>](#),
- [Camilla Pacifici<sup>18</sup>](#),
- [Keren Sharon](#) ORCID: [orcid.org/0000-0002-7559-0864<sup>19</sup>](https://orcid.org/0000-0002-7559-0864) &
- [Francesco Valentino<sup>2,6</sup>](#)

*Nature* volume 597, pages 485–488 (2021)

- 783 Accesses
- 349 Altmetric

- [Metrics details](#)

## Subjects

- [Astrophysical dust](#)
- [Early universe](#)
- [Galaxies and clusters](#)

## Abstract

Star formation in half of massive galaxies was quenched by the time the Universe was 3 billion years old<sup>1</sup>. Very low amounts of molecular gas seem to be responsible for this, at least in some cases<sup>2,3,4,5,6,7</sup>, although morphological gas stabilization, shock heating or activity associated with accretion onto a central supermassive black hole are invoked in other cases<sup>8,9,10,11</sup>. Recent studies of quenching by gas depletion have been based on upper limits that are insufficiently sensitive to determine this robustly<sup>2,3,4,5,6,7</sup>, or stacked emission with its problems of averaging<sup>8,9</sup>. Here we report 1.3 mm observations of dust emission from 6 strongly lensed galaxies where star formation has been quenched, with magnifications of up to a factor of 30. Four of the six galaxies are undetected in dust emission, with an estimated upper limit on the dust mass of 0.0001 times the stellar mass, and by proxy (assuming a Milky Way molecular gas-to-dust ratio) 0.01 times the stellar mass in molecular gas. This is two orders of magnitude less molecular gas per unit stellar mass than seen in star forming galaxies at similar redshifts<sup>12,13,14</sup>. It remains difficult to extrapolate from these small samples, but these observations establish that gas depletion is responsible for a cessation of star formation in some fraction of high-redshift galaxies.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

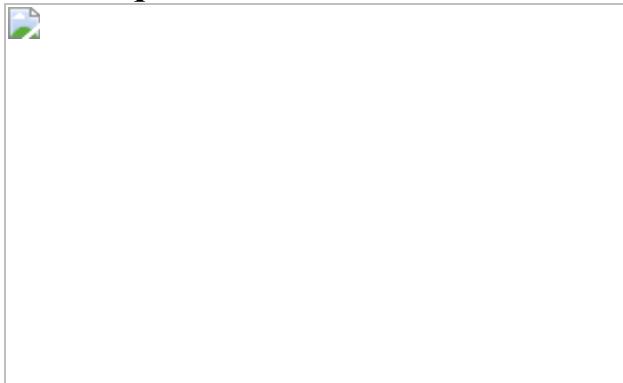
[Rent or Buy](#)

All prices are NET prices.

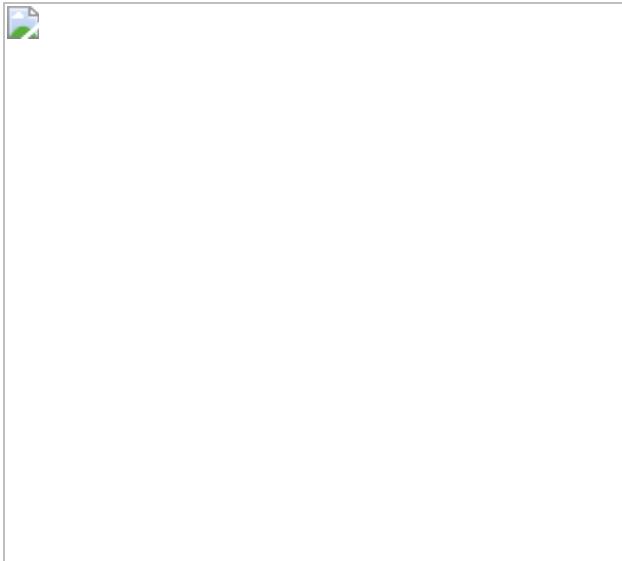
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

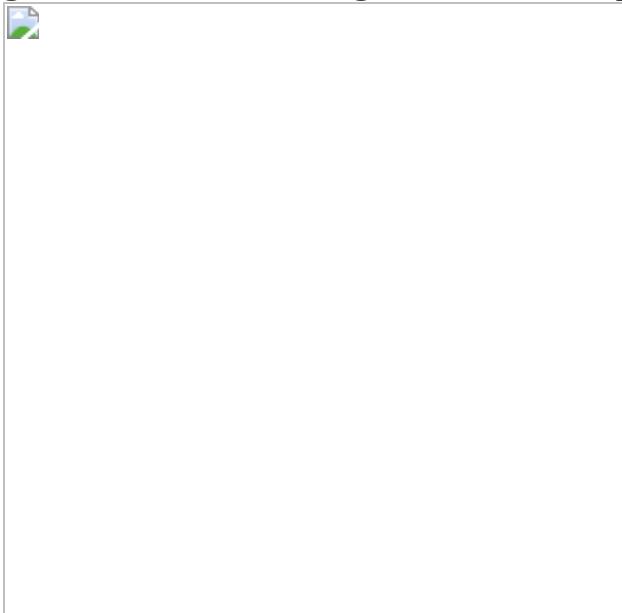
**Fig. 1: Images of six massive lensed galaxies for which star formation has been quenched.**



**Fig. 2: Low dust masses for quenched galaxies.**



**Fig. 3: Low molecular gas masses compared to star forming galaxies.**



## Data availability

Data that support the findings of this study are publicly available through the [ALMA Science Archive](#) under project codes 2018.1.00276.S and 2019.1.00227.S and the Barbara A. Mikulski Archive for Space Telescope under project code [HST-GO-15663](#) (including additional archival data from project codes [HST-GO-9722](#), [HST-GO-9836](#), [HST-SNAP-11103](#), [HST-GO-11591](#), [HST-GO-12099](#), [HST-GO-12100](#), [HST-SNAP-12884](#), [HST-GO-13459](#), [HST-SNAP-14098](#), [HST-GO-14205](#), [HST-GO-14496](#), [HST-SNAP-](#)

[15132](#) and [HST-GO-15466](#)). All HST and ALMA mosaics are publicly available at <https://doi.org/10.5281/zenodo.5009315>. Derived data and codes supporting the findings of this study are available from the corresponding author upon request. [Source data](#) are provided with this paper.

## References

1. 1.

Muzzin, A. et al. The evolution of the stellar mass functions of star-forming and quiescent galaxies to  $z=4$  from the COSMOS/UltraVISTA survey. *Astrophys. J.* **777**, 18 (2013).

2. 2.

Sargent, M. et al. A direct constraint on the gas content of a massive, passively evolving elliptical galaxy at  $z = 1.43$ . *Astrophys. J.* **806**, 20 (2015).

3. 3.

Spilker, J. et al. Molecular gas contents and scaling relations for massive, passive galaxies at intermediate redshifts from the LEGA-C survey. *Astrophys. J.* **860**, 103 (2018).

4. 4.

Bezanson, R. et al. Extremely low molecular gas content in a compact, quiescent galaxy at  $z = 1.522$ . *Astrophys. J.* **873**, 19 (2019).

5. 5.

Zavala, J. et al. On the gas content, star formation efficiency, and environmental quenching of massive galaxies in protoclusters at  $z \sim 2.0\text{--}2.5$ . *Astrophys. J.* **887**, 183 (2019).

6. 6.

Caliendo, J. et al. Early science with the large millimeter telescope: constraining the gas fraction of a compact quiescent galaxy at  $z = 1.883$ . *Astrophys. J. Lett.* **910**, L7 (2021).

7. 7.

Williams, C. et al. ALMA measures rapidly depleted molecular gas reservoirs in massive quiescent galaxies at  $z \sim 1.5$ . *Astrophys. J.* **908**, 54 (2021).

8. 8.

Gobat, R. et al. The unexpectedly large dust and gas content of quiescent galaxies at  $z > 1.4$ . *Nat. Astron.* **2**, 239–246 (2018).

9. 9.

Magdis, G. et al. The interstellar medium of quiescent galaxies and its evolution with time. *Astron. Astrophys.* **647**, 33 (2021).

10. 10.

Suess, K. et al. Massive quenched galaxies at  $z \sim 0.7$  retain large molecular gas reservoirs. *Astrophys. J.* **846**, 14 (2017).

11. 11.

Hayashi, M. et al. Molecular gas reservoirs in cluster galaxies at  $z = 1.46$ . *Astrophys. J.* **856**, 118 (2018).

12. 12.

Tacconi, L. et al. High molecular gas fractions in normal massive star-forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).

13. 13.

Genzel, R. et al. Combined CO and dust scaling relations of depletion time and molecular gas fractions with cosmic time, specific star-

formation rate, and stellar mass. *Astrophys. J.* **800**, 20 (2015).

14. 14.

Tacconi, L. et al. PHIBSS: unified scaling relations of gas depletion time and molecular gas fractions. *Astrophys. J.* **853**, 179 (2018).

15. 15.

Ebeling, H. et al. Thirty-fold: extreme gravitational lensing of a quiescent galaxy at  $z=1.6$ . *Astrophys. J.* **852**, 7 (2018).

16. 16.

Newman, N. et al. Resolving quiescent galaxies at  $z>2$ . I. Search for gravitationally lensed sources and characterization of their structure, stellar populations, and line emission. *Astrophys. J.* **862**, 125 (2018).

17. 17.

Toft, S. et al. A massive, dead disk galaxy in the early Universe. *Nature* **546**, 510–513 (2017).

18. 18.

Man, A. et al. An exquisitely deep view of quenching galaxies through the gravitational lens: Stellar population, morphology, and ionized gas. Preprint at <https://arxiv.org/abs/2106.08338> (2021).

19. 19.

Scoville, N. et al. ISM masses and the star formation law at  $Z = 1$  to 6: ALMA observations of dust continuum in 145 galaxies in the COSMOS survey field. *Astrophys. J.* **820**, 83 (2016).

20. 20.

Tadaki, K. et al. Bulge-forming galaxies with an extended rotating disk at  $z \sim 2$ . *Astrophys. J.* **824**, 175 (2017).

21. 21.

Saintonge, A. et al. xCOLD GASS: the complete IRAM 30 m legacy survey of molecular gas for galaxy evolution studies. *Astrophys. J. Suppl. Ser.* **233**, 22 (2017).

22. 22.

Li, Z. et al. The evolution of the interstellar medium in post-starburst galaxies. *Astrophys. J.* **879**, 131 (2019).

23. 23.

Thomas, D. et al. The epochs of early-type galaxy formation as a function of environment. *Astrophys. J.* **621**, 673 (2005).

24. 24.

Valentino, F. et al. Quiescent galaxies 1.5 billion years after the Big Bang and their progenitors. *Astrophys. J.* **889**, 93 (2020).

25. 25.

Lagos, C. et al. The origin of the atomic and molecular gas contents of early-type galaxies. II. Misaligned gas accretion. *Mon. Notices R. Astron. Soc.* **448**, 1271–1287 (2015).

26. 26.

Dave, R. et al. SIMBA: cosmological simulations with black hole growth and feedback. *Mon. Notices R. Astron. Soc.* **486**, 2827–2849 (2019).

27. 27.

Keres, D. et al. How do galaxies get their gas? *Mon. Notices R. Astron. Soc.* **363**, 2–28 (2005).

28. 28.

Dekel, A. et al. Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).

29. 29.

Whitaker, K. et al. Constraining the low-mass slope of the star formation sequence at  $0.5 < z < 2.5$ . *Astrophys. J.* **775**, 104 (2014).

30. 30.

Ciotti, L. et al. Radiative feedback from massive black holes in elliptical galaxies: AGN flaring and central starburst fueled by recycled gas. *Astrophys. J.* **665**, 1038–1056 (2007).

31. 31.

Akhshik, M. et al. Recent star formation in a massive slowly quenched lensed quiescent galaxy at  $z = 1.88$ . *Astrophys. J. Lett.* **907**, L8 (2021).

32. 32.

Dekel, A. & Birnboim, Y. Galaxy bimodality due to cold flows and shock heating. *Mon. Notices R. Astron. Soc.* **368**, 2–20 (2006).

33. 33.

Cheung, E. et al. Suppressing star formation in quiescent galaxies with supermassive black hole winds. *Nature* **533**, 504–508 (2016).

34. 34.

Whitaker, K. et al. Quiescent galaxies in the 3D-HST survey: spectroscopic confirmation of a large number of galaxies with relatively old stellar populations at  $z \sim 2$ . *Astrophys. J. Lett.* **770**, 39 (2013).

35. 35.

Johansson, P. et al. Gravitational heating helps make massive galaxies red and dead. *Astrophys. J. Lett.* **697**, L38–L43 (2009).

36. 36.

Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pac.* **115**, 763–795 (2003).

37. 37.

Akhshik, M. et al. REQUIEM-2D methodology: spatially resolved stellar populations of massive lensed quiescent galaxies from Hubble Space Telescope 2D grism spectroscopy. *Astrophys. J.* **900**, 184 (2020).

38. 38.

Calzetti, D. et al. The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).

39. 39.

Bruzual, G., & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Notices R. Astron. Soc.* **344**, 1000–1028 (2003).

40. 40.

Lee, B. et al. The intrinsic characteristics of galaxies on the SFR- $M_*$  plane at  $1.2 < z < 4$ . I. The correlation between stellar age, central density, and position relative to the main sequence. *Astrophys. J.* **853**, 131 (2018).

41. 41.

Salmon, B. et al. Breaking the curve with CANDELS: a Bayesian approach to reveal the non-universality of the dust-attenuation law at high redshift. *Astrophys. J.* **827**, 20 (2016).

42. 42.

Salim, S. et al. Dust attenuation curves in the local universe: demographics and new laws for star-forming galaxies and high-redshift analogs. *Astrophys. J.* **859**, 11 (2018).

43. 43.

Leja, J. et al. An older, more quiescent universe from panchromatic SED fitting of the 3D-HST survey. *Astrophys. J.* **877**, 140 (2019).

44. 44.

Conroy, C., Gunn, J., & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699**, 486–506 (2009).

45. 45.

Kriek, M., & Conroy, C. The dust attenuation law in distant galaxies: evidence for variation with spectral type. *Astrophys. J. Lett.* **775**, 16 (2013).

46. 46.

Johansson, D., Sigurdarson, H. & Horellou, C. A LABOCA survey of submillimeter galaxies behind galaxy clusters. *Astron. Astrophys.* **527**, 117 (2011).

47. 47.

Greve, T. et al. Submillimeter observations of millimeter bright galaxies discovered by the South Pole Telescope. *Astrophys. J.* **756**, 101 (2012).

48. 48.

Scoville, N. et al. The evolution of interstellar medium mass probed by dust emission: ALMA observations at  $z = 0.3\text{--}2$ . *Astrophys. J.* **783**, 84 (2014)

49. 49.

Zhang, C. et al. Nearly all massive quiescent disk galaxies have a surprisingly large atomic gas reservoir. *Astrophys. J. Lett.* **884**, 52 (2019).

50. 50.

Sage, L. et al. The cool ISM in elliptical galaxies. I. A survey of molecular gas. *Astrophys. J.* **657**, 232–240 (2007).

51. 51.

Li, Q. et al. The dust-to-gas and dust-to-metal ratio in galaxies from  $z = 0$  to 6. *Mon. Notices R. Astron. Soc.* **490**, 1425–1436 (2019).

52. 52.

Smercina, A. et al. After the fall: the dust and gas in E+A post-starburst galaxies. *Astrophys. J.* **855**, 51 (2018).

53. 53.

Morishita, T. et al. Extremely low molecular gas content in the vicinity of a red nugget galaxy at  $z = 1.91$ . *Astrophys. J.* **908**, 163 (2021).

54. 54.

Smith, M. et al. The Herschel Reference Survey: dust in early-type galaxies and across the Hubble sequence. *Astrophys. J.* **748**, 123 (2012).

55. 55.

Saintonge, A. et al. Validation of the equilibrium model for galaxy evolution to  $z \sim 3$  through molecular gas and dust observations of lensed star-forming galaxies. *Astrophys. J.* **778**, 2 (2013).

56. 56.

Franco, M. et al. GOODS-ALMA: the slow downfall of star formation in  $z = 2$ -3 massive galaxies. *Astron. Astrophys.* **643**, 30 (2020).

57. 57.

Tacconi, L. et al. Submillimeter galaxies at  $z \sim 2$ : evidence for major mergers and constraints on lifetimes, IMF, and CO-H<sub>2</sub> conversion factor. *Astrophys. J.* **680**, 246–262 (2008).

58. 58.

Daddi, E. et al. Very high gas fractions and extended gas reservoirs in  $z = 1.5$  disk galaxies. *Astrophys. J.* **713**, 686–707 (2010).

59. 59.

Silverman, J. et al. A higher efficiency of converting gas to stars pushes galaxies at  $z \sim 1.6$  well above the star-forming main sequence. *Astrophys. J. Lett.* **812**, L23 (2015).

60. 60.

Decarli, R. et al. The ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: molecular gas reservoirs in high-redshift galaxies. *Astrophys. J.* **833**, 70 (2016).

61. 61.

Rudnick, G. et al. Deep CO(1-0) observations of  $z = 1.62$  cluster galaxies with substantial molecular gas reservoirs and normal star formation efficiencies. *Astrophys. J.* **849**, 27 (2017).

62. 62.

Spilker, J. et al. Low gas fractions connect compact star-forming galaxies to their  $z \sim 2$  quiescent descendants. *Astrophys. J.* **832**, 19 (2016).

63. 63.

Aravena, M. et al. The ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: the nature of the faintest dusty star-forming galaxies. *Astrophys. J.* **901**, 79 (2020).

## Acknowledgements

This paper makes use of ADS/JAO.ALMA 2018.1.00276.S and ADS/JAO.ALMA 2019.1.00227.S ALMA data. ALMA is a partnership of the European Southern Observatory (ESO; representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan) and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The NRAO is a facility of the NSF operated under cooperative agreement by Associated Universities. This work uses observations from the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, under NASA contract NAS 5-26555. K.E.W. wishes to acknowledge funding from the Alfred P. Sloan Foundation, HST-GO-14622 and HST-GO-15663. C.C.W. acknowledges support from the NSF Astronomy and Astrophysics Fellowship grant AST-1701546 and from the NIRCam Development Contract NAS50210 from NASA Goddard Space Flight Center to the University of Arizona. S.T. acknowledges support from the ERC Consolidator Grant funding scheme (project ConTExt, grant no. 648179), F.V. from the Carlsberg Foundation Research Grant CF18-0388, and G.E.M. from the Villum Fonden research grant 13160. The Cosmic Dawn Center is funded by the Danish National Research Foundation under grant no. 140. C.P. is supported by the Canadian Space Agency under a contract with NRC Herzberg Astronomy and Astrophysics. M.A. acknowledges support from NASA under award no. 80NSSC19K1418. J.S.S. is a NHFP Hubble Fellow supported by NASA Hubble Fellowship grant no. HF2-51446 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, for NASA, under contract NAS5-26555. A.M. is supported by a Dunlap Fellowship at the Dunlap Institute for Astronomy & Astrophysics, funded through an endowment established by the David Dunlap family and the University of Toronto. D.N. acknowledges support from the NSF via AST-1908137.

# Author information

## Affiliations

1. Department of Astronomy, University of Massachusetts, Amherst, MA, USA

Katherine E. Whitaker & Alexandra Pope

2. Cosmic Dawn Center (DAWN), Copenhagen, Denmark

Katherine E. Whitaker, Sune Toft, Desika Narayanan, Georgios E. Magdis, Gabriel B. Brammer & Francesco Valentino

3. Steward Observatory, University of Arizona, Tucson, AZ, USA

Christina C. Williams

4. Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Toronto, Ontario, Canada

Lamiya Mowla

5. Department of Astronomy, University of Texas at Austin, Austin, TX, USA

Justin S. Spilker

6. Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Sune Toft, Georgios E. Magdis, Gabriel B. Brammer & Francesco Valentino

7. Department of Astronomy, University of Florida, Gainesville, FL, USA

Desika Narayanan

8. DTU-Space, Technical University of Denmark, Kongens Lyngby,  
Denmark

Georgios E. Magdis

9. Astronomy Department, Yale University, New Haven, CT, USA

Pieter G. van Dokkum

10. Department of Physics, University of Connecticut, Storrs, CT, USA

Mohammad Akhshik

11. Department of Physics and Astronomy and PITT PACC, University of  
Pittsburgh, Pittsburgh, PA, USA

Rachel Bezanson

12. Department of Astronomy and Astrophysics, The Pennsylvania State  
University, University Park, PA, USA

Joel Leja

13. Institute for Computational and Data Sciences, The Pennsylvania State  
University, University Park, PA, USA

Joel Leja

14. Institute for Gravitation and the Cosmos, The Pennsylvania State  
University, University Park, PA, USA

Joel Leja

15. Department of Physics & Astronomy, University of British Columbia,  
Vancouver, British Columbia, Canada

Allison Man

16. Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA

Erica J. Nelson

17. Université Lyon, Université Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, Saint-Genis-Laval, France

Johan Richard

18. Space Telescope Science Institute, Baltimore, MD, USA

Camilla Pacifici

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

Keren Sharon

## Contributions

K.E.W. proposed and carried out the observations, conducted the analysis, and wrote the majority of the manuscript. C.C.W. performed the weighted stack of the data, helped to create Figs. 2 and 3, and edited the main text of the manuscript. L.M. performed direct analysis of the ALMA flux densities and created the images in Fig. 1. J.S.S. carried out the reduction and direct analysis of the raw ALMA data. M.A. reduced the HST images, and M.A. and J.L. performed a stellar population synthesis analysis. G.E.M., A.P., S.T. and F.V. helped to interpret the millimetre data and contributed to the dust and gas mass analysis. D.N. helped to interpret the data in the context of cosmological simulation models. All authors, including R.B., G.B.B., J.L., A.M., E.J.N., C.P., K.S. and P.G.v.D., contributed to the overall interpretation of the results and aspects of the analysis and writing.

## Corresponding author

Correspondence to Katherine E. Whitaker.

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Claudia Maraston and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Source data

[\*\*Source Data Fig. 2\*\*](#)

[\*\*Source Data Fig. 3\*\*](#)

## Rights and permissions

[\*\*Reprints and Permissions\*\*](#)

## About this article

## Cite this article

Whitaker, K.E., Williams, C.C., Mowla, L. *et al.* Quenching of star formation from a lack of inflowing gas to galaxies. *Nature* **597**, 485–488 (2021). <https://doi.org/10.1038/s41586-021-03806-7>

- Received: 18 November 2020

- Accepted: 06 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03806-7>

## 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

## Running on fumes

- Claudia Maraston

News & Views 23 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03806-7>

- Article
- [Published: 22 September 2021](#)

# Normal, dust-obscured galaxies in the epoch of reionization

- [Y. Fudamoto](#) ORCID: [orcid.org/0000-0001-7440-8832](https://orcid.org/0000-0001-7440-8832)<sup>1,2,3</sup>,
- [P. A. Oesch](#) ORCID: [orcid.org/0000-0001-5851-6649](https://orcid.org/0000-0001-5851-6649)<sup>1,4</sup>,
- [S. Schouws](#)<sup>5</sup>,
- [M. Stefanon](#) ORCID: [orcid.org/0000-0001-7768-5309](https://orcid.org/0000-0001-7768-5309)<sup>5</sup>,
- [R. Smit](#)<sup>6</sup>,
- [R. J. Bouwens](#)<sup>5</sup>,
- [R. A. A. Bowler](#) ORCID: [orcid.org/0000-0003-3917-1678](https://orcid.org/0000-0003-3917-1678)<sup>7</sup>,
- [R. Endsley](#) ORCID: [orcid.org/0000-0003-4564-2771](https://orcid.org/0000-0003-4564-2771)<sup>8</sup>,
- [V. Gonzalez](#)<sup>9,10</sup>,
- [H. Inami](#) ORCID: [orcid.org/0000-0003-4268-0393](https://orcid.org/0000-0003-4268-0393)<sup>11</sup>,
- [I. Labbe](#)<sup>12</sup>,
- [D. Stark](#)<sup>8</sup>,
- [M. Aravena](#) ORCID: [orcid.org/0000-0002-6290-3198](https://orcid.org/0000-0002-6290-3198)<sup>13</sup>,
- [L. Barrufet](#) ORCID: [orcid.org/0000-0003-1641-6185](https://orcid.org/0000-0003-1641-6185)<sup>1</sup>,
- [E. da Cunha](#)<sup>14,15</sup>,
- [P. Dayal](#)<sup>16</sup>,
- [A. Ferrara](#) ORCID: [orcid.org/0000-0002-9400-7312](https://orcid.org/0000-0002-9400-7312)<sup>17</sup>,
- [L. Graziani](#) ORCID: [orcid.org/0000-0002-9231-1505](https://orcid.org/0000-0002-9231-1505)<sup>18,20,27</sup>,
- [J. Hodge](#)<sup>5</sup>,
- [A. Hutter](#) ORCID: [orcid.org/0000-0003-3760-461X](https://orcid.org/0000-0003-3760-461X)<sup>16</sup>,
- [Y. Li](#)<sup>21,22</sup>,
- [I. De Looze](#)<sup>23,24</sup>,
- [T. Nanayakkara](#) ORCID: [orcid.org/0000-0003-2804-0648](https://orcid.org/0000-0003-2804-0648)<sup>12</sup>,
- [A. Pallottini](#) ORCID: [orcid.org/0000-0002-7129-5761](https://orcid.org/0000-0002-7129-5761)<sup>17</sup>,

- [D. Riechers<sup>25</sup>](#),
- [R. Schneider<sup>18,19,26,27</sup>](#),
- [G. Ucci<sup>16</sup>](#),
- [P. van der Werf<sup>5</sup>](#) &
- [C. White<sup>8</sup>](#)

[\*Nature\*](#) volume **597**, pages 489–492 (2021)

- 627 Accesses
- 237 Altmetric
- [Metrics details](#)

## Subjects

- [Astrophysical dust](#)
- [Early universe](#)
- [Galaxies and clusters](#)
- [Interstellar medium](#)

## Abstract

Over the past decades, rest-frame ultraviolet (UV) observations have provided large samples of UV luminous galaxies at redshift ( $z$ ) greater than 6 (refs. [1,2,3](#)), during the so-called epoch of reionization. While a few of these UV-identified galaxies revealed substantial dust reservoirs<sup>[4,5,6,7](#)</sup>, very heavily dust-obscured sources at these early times have remained elusive. They are limited to a rare population of extreme starburst galaxies<sup>[8,9,10,11,12](#)</sup> and companions of rare quasars<sup>[13,14](#)</sup>. These studies conclude that the contribution of dust-obscured galaxies to the cosmic star formation rate density at  $z > 6$  is sub-dominant. Recent ALMA and Spitzer observations have identified a more abundant, less extreme population of obscured galaxies at  $z = 3\text{--}6$  (refs. [15,16](#)). However, this population has not been confirmed in the reionization epoch so far. Here, we report the discovery of two dust-obscured star-forming galaxies at  $z = 6.6813 \pm 0.0005$  and  $z =$

$7.3521 \pm 0.0005$ . These objects are not detected in existing rest-frame UV data and were discovered only through their far-infrared [C ii] lines and dust continuum emission as companions to typical UV-luminous galaxies at the same redshift. The two galaxies exhibit lower infrared luminosities and star-formation rates than extreme starbursts, in line with typical star-forming galaxies at  $z \approx 7$ . This population of heavily dust-obscured galaxies appears to contribute 10–25% to the  $z > 6$  cosmic star formation rate density.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

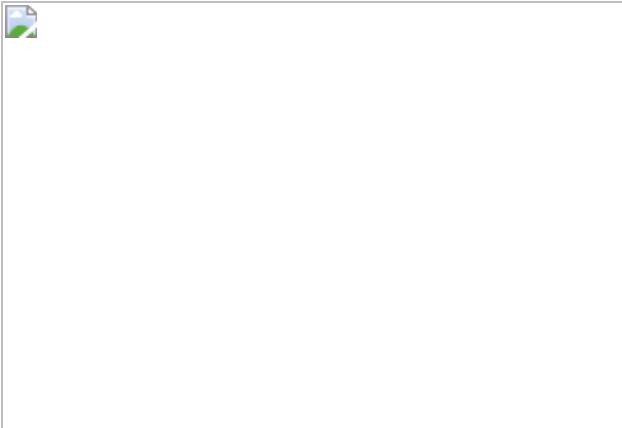
All prices are NET prices.

### Additional access options:

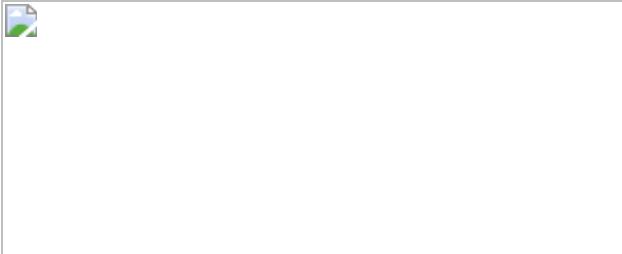
- [Log in](#)
- [Access through your institution](#)

- [Learn about institutional subscriptions](#)

**Fig. 1: [C ii] 158  $\mu\text{m}$  line and dust emission detections.**



**Fig. 2: Estimated properties of REBELS-29-2 and REBELS-12-2.**



**Fig. 3: Contribution of obscured galaxies to the cosmic SFR density \(\langle \rho\_{\text{SFR}} \rangle\).**



## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. This paper

makes use of the following ALMA data: ADS/JAO.ALMA #2019.1.01634.L.

## Code availability

The codes used to reduce and analyse the ALMA data are publicly available. The code used to model the optical-to-infrared SEDs is accessible through GitHub (<https://github.com/ACCarnall/bagpipes>).

## References

1. 1.  
Madau, P. & Dickinson, M. Cosmic star-formation history. *Annu. Rev. Astron. Astrophys.* **52**, 415–486 (2014).
2. 2.  
Bouwens, R. J. et al. UV luminosity functions at redshifts  $z \sim 4$  to  $z \sim 10$ : 10,000 galaxies from HST legacy fields. *Astrophys. J.* **803**, 34 (2015).
3. 3.  
Ono, Y. et al. Great optically luminous dropout research using subaru HSC (GOLDRUSH). I. UV luminosity functions at  $z \sim 4\text{--}7$  derived with the half-million dropouts on the 100 deg<sup>2</sup> sky. *Publ. Astron. Soc. Jpn.* **70**, S10 (2018).
4. 4.  
Watson, D. et al. A dusty, normal galaxy in the epoch of reionization. *Nature* **519**, 327–330 (2015).
5. 5.

Hashimoto, T. et al. Big three dragons: a  $z = 7.15$  Lyman-break galaxy detected in [O iii] 88  $\mu\text{m}$ , [C ii] 158  $\mu\text{m}$ , and dust continuum with ALMA. *Publ. Astron. Soc. Jpn.* **71**, 71 (2019).

6. 6.

Tamura, Y. et al. Detection of the far-infrared [O iii] and dust emission in a galaxy at redshift 8.312: early metal enrichment in the heart of the reionization era. *Astrophys. J.* **874**, 27 (2019).

7. 7.

Bakx, T. J. L. C. et al. ALMA uncovers the [C ii] emission and warm dust continuum in a  $z = 8.31$  Lyman break galaxy. *Mon. Not. R. Astron. Soc.* **493**, 4294–4307 (2020).

8. 8.

Riechers, D. A. et al. A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. *Nature* **496**, 329–333 (2013).

9. 9.

Strandet, M. L. et al. ISM properties of a massive dusty star-forming galaxy discovered at  $z \sim 7$ . *Astrophys. J. Lett.* **842**, L15 (2017).

10. 10.

Marrone, D. P. et al. Galaxy growth in a massive halo in the first billion years of cosmic history. *Nature* **553**, 51–54 (2018).

11. 11.

Dudzevičiūtė, U. et al. An ALMA survey of the SCUBA-2 CLS UDS field: physical properties of 707 sub-millimetre galaxies. *Mon. Not. R. Astron. Soc.* **494**, 3828–3860 (2020).

12. 12.

Riechers, D. A. et al. COLDz: a high space density of massive dusty starburst galaxies  $\sim 1$  billion years after the big bang. *Astrophys. J.* **895**, 81 (2020).

13. 13.

Decarli, R. et al. Rapidly star-forming galaxies adjacent to quasars at redshifts exceeding 6. *Nature* **545**, 457–461 (2017).

14. 14.

Mazzucchelli, C. et al. Spectral energy distributions of companion galaxies to  $z \sim 6$  quasars. *Astrophys. J.* **881**, 163 (2019).

15. 15.

Wang, T. et al. A dominant population of optically invisible massive galaxies in the early Universe. *Nature* **572**, 211–214 (2019).

16. 16.

Williams, C. C. et al. Discovery of a dark, massive, ALMA-only galaxy at  $z \sim 5\text{--}6$  in a tiny 3 mm survey. *Astrophys. J.* **884**, 154 (2019).

17. 17.

Bouwens, R. J. et al. Reionization era bright emission line survey: selection and characterization of luminous interstellar medium reservoirs in the  $z > 6.5$  Universe. Preprint at <https://arxiv.org/abs/2106.13719> (2021).

18. 18.

Bowler, R. A. A. et al. Obscured star formation in bright  $z = 7$  Lyman-break galaxies. *Mon. Not. R. Astron. Soc.* **481**, 1631–1644 (2018).

19. 19.

Schreiber, C. et al. The Herschel view of the dominant mode of galaxy growth from  $z=4$  to the present day. *Astron. Astrophys.* **575**, A74 (2015).

20. 20.

Swinbank, A. M. et al. An ALMA survey of sub-millimetre galaxies in the Extended Chandra Deep Field South: the far-infrared properties of SMGs. *Mon. Not. R. Astron. Soc.* **438**, 1267–1287 (2014).

21. 21.

Walter, F. et al. The intense starburst HDF 850.1 in a galaxy overdensity at  $z \approx 5.2$  in the Hubble Deep Field. *Nature* **486**, 233–236 (2012).

22. 22.

Casey, C. M. et al. The brightest galaxies in the dark ages: galaxies' dust continuum emission during the reionization era. *Astrophys. J.* **862**, 77 (2018).

23. 23.

Zavala, J. A. et al. The evolution of the IR luminosity function and dust-obscured star formation over the past 13 billion years. *Astrophys. J.* **909**, 165 (2021).

24. 24.

McCracken, H. J. et al. UltraVISTA: a new ultra-deep near-infrared survey in COSMOS. *Astron. Astrophys.* **544**, A156 (2012).

25. 25.

Jarvis, M. J. et al. The VISTA Deep Extragalactic Observations (VIDEO) survey. *Mon. Not. R. Astron. Soc.* **428**, 1281–1295 (2013).

26. 26.

Erben, T. et al. CARS: the CFHTLS-Archive-Research Survey. I. Five-band multi-colour data from 37 sq. deg. CFHTLS-wide observations. *Astron. Astrophys.* **493**, 1197–1222 (2009).

27. 27.

Aihara, H. et al. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publ. Astron. Soc. Jpn.* **70**, S8 (2018).

28. 28.

Bowler, R. A. A. et al. A lack of evolution in the very bright end of the galaxy luminosity function from  $z = 8$  to 10. *Mon. Not. R. Astron. Soc.* **493**, 2059–2084 (2020).

29. 29.

Stefanon, M. et al. The brightest  $z \gtrsim 8$  galaxies over the COSMOS UltraVISTA field. *Astrophys. J.* **883**, 99 (2019).

30. 30.

Bowler, R. A. A., Dunlop, J. S., McLure, R. J. & McLeod, D. J. Unveiling the nature of bright  $z = 7$  galaxies with the Hubble Space Telescope. *Mon. Not. R. Astron. Soc.* **466**, 3612–3635 (2017).

31. 31.

Schaerer, D. et al. The ALPINE-ALMA [C ii] survey. Little to no evolution in the [C ii]-SFR relation over the last 13 Gyr. *Astron. Astrophys.* **643**, A3 (2020).

32. 32.

De Looze, I. et al. The applicability of far-infrared fine-structure lines as star formation rate tracers over wide ranges of metallicities and galaxy types. *Astron. Astrophys.* **568**, A62 (2014).

33. 33.

Carnall, A. C., McLure, R. J., Dunlop, J. S. & Davé, R. Inferring the star formation histories of massive quiescent galaxies with BAGPIPES: evidence for multiple quenching mechanisms. *Mon. Not. R. Astron. Soc.* **480**, 4379–4401 (2018).

34. 34.

Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).

35. 35.

Kroupa, P. & Boily, C. M. On the mass function of star clusters. *Mon. Not. R. Astron. Soc.* **336**, 1188–1194 (2002).

36. 36.

Byler, N., Dalcanton, J. J., Conroy, C. & Johnson, B. D. Nebular continuum and line emission in stellar population synthesis models. *Astrophys. J.* **840**, 44 (2017).

37. 37.

Ferland, G. J. et al. The 2017 release Cloudy. *Rev. Mexic. Astron. Astrof.* **53**, 385–438 (2017).

38. 38.

Calzetti, D. et al. The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).

39. 39.

Charlot, S. & Fall, S. M. A simple model for the absorption of starlight by dust in galaxies. *Astrophys. J.* **539**, 718–731 (2000).

40. 40.

Draine, B. T. & Li, A. Infrared emission from interstellar dust. IV. The silicate-graphite-PAH model in the post-Spitzer era. *Astrophys. J.* **657**, 810–837 (2007).

41. 41.

Wang, R. et al. Star formation and gas kinematics of quasar host galaxies at  $z \sim 6$ : new insights from ALMA. *Astrophys. J.* **773**, 44 (2013).

42. 42.

Capak, P. L. et al. Galaxies at redshifts 5 to 6 with systematically low dust content and high [C ii] emission. *Nature* **522**, 455–458 (2015).

43. 43.

Dessauges-Zavadsky, M. et al. The ALPINE-ALMA [C ii] survey. Molecular gas budget in the early Universe as traced by [C ii]. *Astron. Astrophys.* **643**, A5 (2020).

44. 44.

Casey, C. M. Far-infrared spectral energy distribution fitting for galaxies near and far. *Mon. Not. R. Astron. Soc.* **425**, 3094–3103 (2012).

45. 45.

Schreiber, C. et al. Dust temperature and mid-to-total infrared color distributions for star-forming galaxies at  $0 < z < 4$ . *Astron. Astrophys.* **609**, A30 (2018).

46. 46.

Faisst, A. L. et al. ALMA characterises the dust temperature of  $z \sim 5.5$  star-forming galaxies. *Mon. Not. R. Astron. Soc.* **498**, 4192–4204 (2020).

47. 47.

da Cunha, E. et al. On the effect of the cosmic microwave background in high-redshift (sub-)millimeter observations. *Astrophys. J.* **766**, 13 (2013).

48. 48.

Laporte, N. et al. Dust in the reionization era: ALMA observations of a  $z = 8.38$  gravitationally lensed galaxy. *Astrophys. J. Lett.* **837**, L21 (2017).

49. 49.

Behrens, C. et al. Dusty galaxies in the epoch of reionization: simulations. *Mon. Not. R. Astron. Soc.* **477**, 552–565 (2018).

50. 50.

Liang, L. et al. On the dust temperatures of high-redshift galaxies. *Mon. Not. R. Astron. Soc.* **489**, 1397–1422 (2019).

51. 51.

Sommovigo, L. et al. Warm dust in high- $z$  galaxies: origin and implications. *Mon. Not. R. Astron. Soc.* **497**, 956–968 (2020).

52. 52.

De Vis, P. et al. A systematic metallicity study of DustPedia galaxies reveals evolution in the dust-to-metal ratios. *Astron. Astrophys.* **623**, A5 (2019).

53. 53.

Mancini, M. et al. Interpreting the evolution of galaxy colours from  $z = 8$  to 5. *Mon. Not. R. Astron. Soc.* **462**, 3130–3145 (2016).

54. 54.

Graziani, L. et al. The assembly of dusty galaxies at  $z \geq 4$ : statistical properties. *Mon. Not. R. Astron. Soc.* **494**, 1071–1088 (2020).

55. 55.

Gruppioni, C. et al. The Herschel PEP/HerMES luminosity function - I. Probing the evolution of PACS selected galaxies to  $z = 4$ . *Mon. Not. R. Astron. Soc.* **432**, 23–52 (2013).

56. 56.

Carilli, C. L. & Walter, F. Cool gas in high-redshift galaxies. *Annu. Rev. Astron. Astrophys.* **51**, 105–161 (2013).

57. 57.

Peebles, P. J. E. *The Large-Scale Structure of the Universe* (Princeton Univ. Press, 1980).

58. 58.

Barone-Nugent, R. L. et al. Measurement of galaxy clustering at  $z \sim 7.2$  and the evolution of galaxy bias from  $3.8 \sim z \sim 8$  in the XDF, GOODS-S, and GOODS-N. *Astrophys. J.* **793**, 17 (2014).

59. 59.

Adelberger, K. L. et al. The spatial clustering of star-forming galaxies at redshifts  $1.4 < z < 3.5$ . *Astrophys. J.* **619**, 697–713 (2005).

60. 60.

Qiu, Y. et al. Dependence of galaxy clustering on UV luminosity and stellar mass at  $z \sim 4–7$ . *Mon. Not. R. Astron. Soc.* **481**, 4885–4894 (2018).

61. 61.

Bhowmick, A. K. et al. Cosmic variance of  $z > 7$  galaxies: prediction from BLUETIDES. *Mon. Not. R. Astron. Soc.* **496**, 754–766 (2020).

62. 62.

Uzgil, B. D. et al. The ALMA spectroscopic survey in the HUDF: a search for [C ii] emitters at  $6 \leq z \leq 8$ . *Astrophys. J.* **912**, 67 (2021).

63. 63.

Whitaker, K. E. et al. The constant average relationship between dust-obscured star formation and stellar mass from  $z = 0$  to  $z = 2.5$ . *Astrophys. J.* **850**, 208 (2017).

64. 64.

Fudamoto, Y. et al. The ALPINE-ALMA [Cii] survey. Dust attenuation properties and obscured star formation at  $z \sim 4.4\text{--}5.8$ . *Astron. Astrophys.* **643**, A4 (2020).

65. 65.

Béthermin, M. et al. Evolution of the dust emission of massive galaxies up to  $z = 4$  and constraints on their dominant mode of star formation. *Astron. Astrophys.* **573**, A113 (2015).

66. 66.

Scoville, N. et al. COSMOS: Hubble Space Telescope observations. *Astrophys. J. Suppl.* **172**, 38–45 (2007).

## Acknowledgements

The authors thank C. Williams for helpful discussions. We acknowledge support from: the Swiss National Science Foundation through the SNSF Professorship grant 190079 (Y.F., P.A.O., L.B.); NAOJ ALMA Scientific Research Grant 2020-16B (Y.F.); TOP grant TOP1.16.057 (RJB, MS); the Nederlandse Onderzoekschool voor Astronomie (S.S.); STFC Ernest

Rutherford Fellowship ST/S004831/1 (R. Smit) and ST/T003596/1 (R.B.); JSPS KAKENHI JP19K23462 and JP21H01129 (HI); European Research Council's starting grant ERC StG-717001 (P.D., A.H., G.U.); the NWO's VIDI grant 016.vidi.189.162 and the European Commission's and University of Groningen's CO-FUND Rosalind Franklin program (P.D.); the Amaldi Research Center funded by the MIUR program "Dipartimento di Eccellenza" CUP:B81I18001170001 (L.G., R. Schneider); the National Science Foundation MRI-1626251 (Y.L.); FONDECYT grant 1211951, "CONICYT+PCI+INSTITUTO MAX PLANCK DE ASTRONOMIA MPG190030" and "CONICYT+PCI+REDES 190194" (M.A.); ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) CE170100013 (E.d.C.); Australian Research Council Laureate Fellowship FL180100060 (T.N.); the ERC Advanced Grant INTERSTELLAR H2020/740120 (A.P., A.F.) and the Carl Friedrich von Siemens-Forschungspreis der Alexander von Humboldt-Stiftung Research Award (A.F.); the VIDI research program 639.042.611 (J.H.); JWST/NIRCam contract to the University of Arizona, NAS5-02015 (R.E.); ERC starting grant 851622 (IDL); the National Science Foundation under grant numbers AST-1614213, AST-1910107, and the Alexander von Humboldt Foundation through a Humboldt Research Fellowship for Experienced Researchers (D.R.). The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant no. 140. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

## Author information

### Affiliations

1. Department of Astronomy, University of Geneva, Versoix, Switzerland  
Y. Fudamoto, P. A. Oesch & L. Barrufet
2. Research Institute for Science and Engineering, Waseda University, Shinjuku, Tokyo, Japan

Y. Fudamoto

3. National Astronomical Observatory of Japan, Mitaka, Tokyo, Japan

Y. Fudamoto

4. Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, København, Denmark

P. A. Oesch

5. Leiden Observatory, Leiden University, Leiden, the Netherlands

S. Schouws, M. Stefanon, R. J. Bouwens, J. Hodge & P. van der Werf

6. Astrophysics Research Institute, Liverpool John Moores University, Liverpool, UK

R. Smit

7. Sub-department of Astrophysics, The Denys Wilkinson Building, University of Oxford, Oxford, UK

R. A. A. Bowler

8. Steward Observatory, University of Arizona, Tucson, AZ, USA

R. Endsley, D. Stark & C. White

9. Departamento de Astronomia, Universidad de Chile, Santiago, Chile

V. Gonzalez

10. Centro de Astrofisica y Tecnologias Afines (CATA), Santiago, Chile

V. Gonzalez

11. Hiroshima Astrophysical Science Center, Hiroshima University, Hiroshima, Japan

H. Inami

12. Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria, Australia

I. Labbe & T. Nanayakkara

13. Nucleo de Astronomia, Facultad de Ingenieria y Ciencias, Universidad Diego Portales, Santiago, Chile

M. Aravena

14. Centre for Radio Astronomy Research, University of Western Australia, Crawley, Western Australia, Australia

E. da Cunha

15. ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Canberra, Australian Capital Territory, Australia

E. da Cunha

16. Kapteyn Astronomical Institute, University of Groningen, Groningen, the Netherlands

P. Dayal, A. Hutter & G. Ucci

17. Scuola Normale Superiore, Pisa, Italy

A. Ferrara & A. Pallottini

18. Dipartimento di Fisica, Sapienza, Universita di Roma, Roma, Italy

L. Graziani & R. Schneider

19. INAF/Osservatorio Astronomico di Roma, Roma, Italy

R. Schneider

20. INAF/Osservatorio Astrofisico di Arcetri, Firenze, Italy

L. Graziani

21. Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA, USA

Y. Li

22. Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA, USA

Y. Li

23. Sterrenkundig Observatorium, Ghent University, Gent, Belgium

I. De Looze

24. Dept. of Physics & Astronomy, University College London, London, UK

I. De Looze

25. Cornell University, Ithaca, NY, USA

D. Riechers

26. Sapienza School for Advanced Studies, Roma, Italy

R. Schneider

27. INFN, Roma, Italy

L. Graziani & R. Schneider

## Contributions

Y.F. wrote the main part of the text, analysed the data and produced most of the figures. P.A.O. contributed text and led the SED fitting and data

analysis. S.S. calibrated the ALMA data and produced images. M.S. performed detailed photometric measurements from the ground-based images. R.S. contributed comparison plots of different galaxy samples. All co-authors contributed to the successful execution of the ALMA program, to the scientific interpretation of the results, and helped to write up this manuscript.

## Corresponding author

Correspondence to Y. Fudamoto.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Marcel Neeleman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Optical/NIR images and full SEDs of the UV-luminous targets REBELS-29 and REBELS-12.](#)

The cutouts show images from which photometry was extracted. SED fits (bottom-right panels) are performed using the BAGPIPES<sup>33</sup>. In **b** and **d**, blue solid lines and bands represent the median posterior SEDs together with their 68% confidence contours for REBELS-29 and REBELS-12,

respectively. Error bars corresponds to  $1\sigma$  uncertainties, and downward arrows show  $2\sigma$  upper limits. **a** and **c** show that the [C ii] 158  $\mu\text{m}$  emission line redshifts (red) are in perfect agreement with the photometric redshift probability distributions (blue), that had been previously estimated from the optical/NIR photometry for both sources. This confirms their high-redshift nature.

### Extended Data Fig. 2 Optical/NIR/FIR cutouts of the dusty sources REBELS-29-2 and REBELS-12-2.

$\langle (6.5)^{\prime\prime} \rangle \times \langle (6.5)^{\prime\prime} \rangle$  cutouts show the existing ground- and space-based observations: Subaru Hyper Suprime Cam, VISTA VIRCAM, Spitzer IRAC, in addition to the ALMA dust continuum images and continuum subtracted [C ii] 158  $\mu\text{m}$  moment-0 images. White contours show  $(+2,+3,+4,+5)\sigma$  (solid contour) and  $(-5,-4,-3,-2)\sigma$  (dashed contour), if present. A faint low-surface brightness foreground neighbour can be seen  $\sim 2.0''$  to the SE of REBELS-29-2. However, the photometric redshift of this foreground source is  $z_{\rm rm{ph}} = 2.46 \pm 0.07 \pm 0.08$ , and the line frequency of REBELS-29-2 is not consistent with bright FIR emission lines (for example, CO lines) from this foreground redshift. No optical counterparts are found at the location of the ALMA [C ii] and dust continuum positions for both REBELS-29-2 and REBELS-12-2.

### Extended Data Fig. 3 Probing a new parameter space of DSFGs.

**a**, The stellar mass as a function of redshift for DSFGs from the literature. IRAC-selected, H-dropout galaxies (light-grey dots with  $1\sigma$  errorbars<sup>15</sup>) are generally more massive than the two serendipitously detected REBELS galaxies (red dots). Additionally, the redshifts of H-dropouts are extremely uncertain (photo-z). The extremely star-bursting SMG population only shows a small tail of rare sources at  $z > 4$  (shown by dark dots<sup>11</sup>). The blue squares show all the previously known DSFGs at  $z > 5.5$  with spectroscopically measured redshifts, while purple squares correspond to  $z \approx 6$  QSO companion galaxies<sup>13</sup>. These are more extreme sources than

REBELS-12-2 and REBELS-29-2. **b**, The infrared luminosity/SFR<sub>IR</sub> as a function of redshift for the same galaxy samples as on the left. The infrared luminosities and hence SFRs of the newly identified galaxies are substantially lower than typical SMGs at these redshifts. For both panels, error bars correspond to  $1\sigma$  uncertainties, and arrows show  $2\sigma$  upper/lower limits.

### Extended Data Fig. 4 Fraction of obscured star-formation as a function of stellar mass.

The fraction of obscured star-formation,  $\langle f \rangle_{\rm obs} = \langle \rm SFR_{IR} \rangle / (\langle \rm SFR \rangle + \langle \rm UV \rangle)$ , of REBELS-29-2 and REBELS-12-2 (dark coloured squares) is significantly higher than for typical LBGs at their stellar mass. The line shows the observed, constant relation between  $z \approx 0$  and  $z \approx 2.5$  (ref. <sup>63</sup>) assuming a given set of SED templates from Bethermin and colleagues<sup>65</sup>. Blue and brown small points with error bars show stacked results of star-forming galaxies at  $z \approx 4.5$  and at  $z \approx 5.5$ , respectively<sup>64</sup>. The star-formation of extreme starburst galaxies at  $z \approx 5.7$ –6.9 is essentially 100% obscured (SMGs; <sup>12</sup> green small points). The highly obscured star-forming galaxies found as companions of high-redshift quasars at  $z > 6$  (refs. <sup>13,14</sup>) (yellow diamonds) are substantially more massive than the galaxies identified here, as estimated from their dynamical masses. Squares show the obscured fraction of our UV-bright and dusty galaxies. Error bars correspond to  $1\sigma$  uncertainty, and arrows show  $2\sigma$  lower/upper limits. Our discovery of lower mass, obscured galaxies shows that  $f_{\rm obs}$  is likely to vary much more strongly at a fixed stellar mass than previously estimated even in the epoch of reionization.

**Extended Data Table 1 FIR properties observed by ALMA**

**Extended Data Table 2 NIR photometric data**

**Extended Data Table 3 Priors used for panchromatic SED modelling**

## Supplementary information

## Peer Review File

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Fudamoto, Y., Oesch, P.A., Schouws, S. *et al.* Normal, dust-obscured galaxies in the epoch of reionization. *Nature* **597**, 489–492 (2021).  
<https://doi.org/10.1038/s41586-021-03846-z>

- Received: 06 August 2020
- Accepted: 22 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03846-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/s41586-021-03846-z>

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

- Article
- [Published: 22 September 2021](#)

# Single-photon nonlinearity at room temperature

- [Anton V. Zasedatelev](#)<sup>1,2</sup>,
- [Anton V. Baranikov](#)<sup>1,2</sup>,
- [Denis Sannikov](#)<sup>1,2</sup>,
- [Darius Urbonas](#) [ORCID: orcid.org/0000-0001-9919-9221](#)<sup>3</sup>,
- [Fabio Scafirimuto](#) [ORCID: orcid.org/0000-0002-9707-3526](#)<sup>3</sup>,
- [Vladislav Yu. Shishkov](#)<sup>1,2,4,5</sup>,
- [Evgeny S. Andrianov](#)<sup>1,2,4,5</sup>,
- [Yuri E. Lozovik](#)<sup>1,2,4,6,7</sup>,
- [Ullrich Scherf](#) [ORCID: orcid.org/0000-0001-8368-4919](#)<sup>8</sup>,
- [Thilo Stöferle](#) [ORCID: orcid.org/0000-0003-0612-7195](#)<sup>3</sup>,
- [Rainer F. Mahrt](#) [ORCID: orcid.org/0000-0002-9772-1490](#)<sup>3</sup> &
- [Pavlos G. Lagoudakis](#) [ORCID: orcid.org/0000-0002-3557-5299](#)<sup>1,2,9</sup>

[Nature](#) volume **597**, pages 493–497 (2021)

- 3541 Accesses
- 118 Altmetric
- [Metrics details](#)

## Subjects

- [Bose–Einstein condensates](#)
- [Photonic devices](#)
- [Single photons and quantum effects](#)

# Abstract

The recent progress in nanotechnology<sup>1,2</sup> and single-molecule spectroscopy<sup>3,4,5</sup> paves the way for emergent cost-effective organic quantum optical technologies with potential applications in useful devices operating at ambient conditions. We harness a  $\pi$ -conjugated ladder-type polymer strongly coupled to a microcavity forming hybrid light–matter states, so-called exciton-polaritons, to create exciton-polariton condensates with quantum fluid properties. Obeying Bose statistics, exciton-polaritons exhibit an extreme nonlinearity when undergoing bosonic stimulation<sup>6</sup>, which we have managed to trigger at the single-photon level, thereby providing an efficient way for all-optical ultrafast control over the macroscopic condensate wavefunction. Here, we utilize stable excitons dressed with high-energy molecular vibrations, allowing for single-photon nonlinear operation at ambient conditions. This opens new horizons for practical implementations like sub-picosecond switching, amplification and all-optical logic at the fundamental quantum limit.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

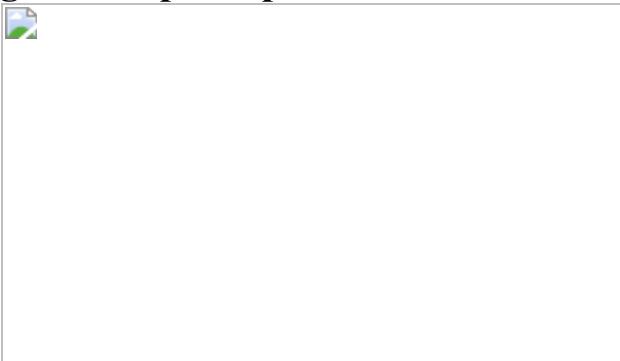
[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

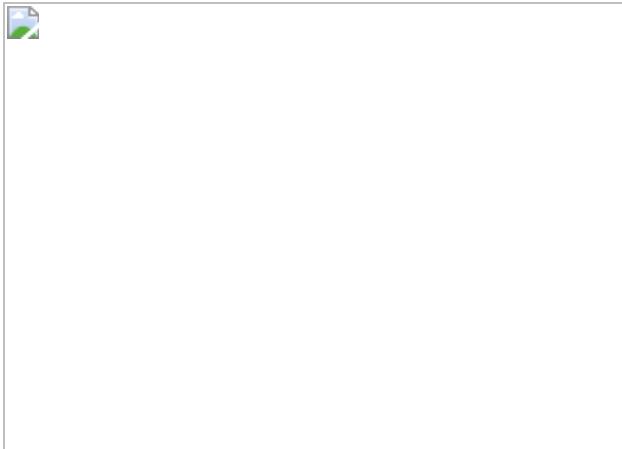
**Fig. 1: The principle of the extreme nonlinearity in organics.**



**Fig. 2: Attojoule polariton switch.**



**Fig. 3: Polariton switching contrast towards the single-photon level.**



**Fig. 4: Single-photon switching for single-shot condensate realizations.**



## Data availability

All data supporting this study are openly available from the University of Southampton repository at <https://doi.org/10.5258/SOTON/D1374>.

## References

1. 1.

Chikkaraddy, R. et al. Single-molecule strong coupling at room temperature in plasmonic nanocavities. *Nature* **535**, 127–130 (2016).

2. 2.

Hail, C. U. et al. Nanoprinting organic molecules at the quantum level. *Nat. Commun.* **10**, 1880 (2019).

3. 3.

Maser, A., Gmeiner, B., Utikal, T., Götzinger, S. & Sandoghdar, V. Few-photon coherent nonlinear optics with a single molecule. *Nat. Photon.* **10**, 450–453 (2016).

4. 4.

Wang, D. et al. Coherent coupling of a single molecule to a scanning Fabry-Perot microcavity. *Phys. Rev. X* **7**, 021014 (2017).

5. 5.

Wang, D. et al. Turning a molecule into a coherent two-level quantum system. *Nat. Phys.* **15**, 483–489 (2019).

6. 6.

Zasedatelev, A. V. et al. A room-temperature organic polariton transistor. *Nat. Photon.* **13**, 378–383 (2019).

7. 7.

Walmsley, I. A. Quantum optics: science and technology in a new light. *Science* **348**, 525–530 (2015).

8. 8.

Chang, D. E., Vučetić, V. & Lukin, M. D. Quantum nonlinear optics—photon by photon. *Nat. Photon.* **6**, 685–694 (2014).

9. 9.

Reiserer, A., Ritter, S. & Rempe, G. Nondestructive detection of an optical photon. *Science* **342**, 1349–1351 (2013).

10. 10.

Shomroni, I. et al. All-optical routing of single photons by a one-atom switch controlled by a single photon. *Science* **346**, 903–906 (2014).

11. 11.

Tiecke, T. G. et al. Nanophotonic quantum phase switch with a single atom. *Nature* **508**, 241–244 (2014).

12. 12.

Hacker, B., Welte, S., Rempe, S. & Ritter, S. A photon–photon quantum gate based on a single atom in an optical resonator. *Nature* **536**, 193–196 (2016).

13. 13.

Volz, T. et al. Ultrafast all-optical switching by single photons. *Nat. Photon.* **6**, 605–609 (2012).

14. 14.

Giesz, V. et al. Coherent manipulation of a solid-state artificial atom with few photons. *Nat. Commun.* **7**, 11986 (2016).

15. 15.

Sun, S., Kim, H., Luo, Z., Solomon, G. S. & Waks, E. A single-photon switch and transistor enabled by a solid-state quantum memory. *Science* **361**, 57–60 (2018).

16. 16.

Dietrich, C. P., Fiore, A., Thompson, M. G., Kamp, M. & Höfling, S. GaAs integrated quantum photonics: towards compact and multi-functional quantum photonic integrated circuits. *Laser Photon. Rev.* **10**, 870–894 (2016).

17. 17.

Peyronel, T. et al. Quantum nonlinear optics with single photons enabled by strongly interacting atoms. *Nature* **488**, 57–60 (2012).

18. 18.

Chen, W. et al. All-optical switch and transistor gated by one stored photon. *Science* **341**, 768–770 (2013).

19. 19.

Gorniaczyk, H., Tresp, C., Schmidt, J., Fedder, H. & Hofferberth, S. Single-photon transistor mediated by interstate Rydberg interactions. *Phys. Rev. Lett.* **113**, 053601 (2014).

20. 20.

Sanvitto, D. & Kéna-Cohen, S. The road towards polaritonic devices. *Nat. Mater.* **15**, 1061–1073 (2016).

21. 21.

Deng, H., Haug, H. & Yamamoto, Y. Exciton-polariton Bose-Einstein condensation. *Rev. Mod. Phys.* **82**, 1489–1537 (2010).

22. 22.

Kasprzak, J. et al. Bose–Einstein condensation of exciton polaritons. *Nature* **443**, 409–414 (2006).

23. 23.

Plumhof, J. D., Stöferle, T., Mai, L., Scherf, U. & Mahrt, R. F. Room-temperature Bose–Einstein condensation of cavity exciton–polaritons in a polymer. *Nat. Mater.* **13**, 247–252 (2014).

24. 24.

Carusotto, I. & Ciuti, C. Quantum fluids of light. *Rev. Mod. Phys.* **85**, 299–366 (2013).

25. 25.

Lerario, G. et al. Room-temperature superfluidity in a polariton condensate. *Nat. Phys.* **13**, 837–841 (2017).

26. 26.

Sun, Z. & Snoke, D. W. Optical switching with organics. *Nat. Photon.* **13**, 370–371 (2019).

27. 27.

Baranikov, A. V. et al. All-optical cascadable universal logic gate with sub-picosecond operation. Preprint at <https://arxiv.org/abs/2005.04802> (2020).

28. 28.

Tartakovskii, A. I. et al. Raman scattering in strongly coupled organic semiconductor microcavities. *Phys. Rev. B* **63**, 121302 (2001).

29. 29.

Coles, D. M. et al. vibrationally assisted polariton-relaxation processes in strongly coupled organic-semiconductor microcavities. *Adv. Funct. Mater.* **21**, 3691–3696 (2011).

30. 30.

Grant, R. T. et al. Efficient radiative pumping of polaritons in a strongly coupled microcavity by a fluorescent molecular dye. *Adv. Opt. Mater.* **4**, 1615–1623 (2016).

31. 31.

Daskalakis, K. S., Maier, S. A. & Kéna-Cohen, S. Spatial coherence and stability in a disordered organic polariton condensate. *Phys. Rev. Lett.* **115**, 035301 (2015).

32. 32.

Bobrovska, N. et al. Dynamical instability of a nonequilibrium exciton-polariton condensate. *ACS Photon.* **5**, 111–118 (2018).

33. 33.

Scherf, U., Bohnen, A. & Müllen, K. Polyarylenes and poly(arylenevinylene)s, 9 The oxidized states of a (1,4-phenylene) ladder polymer. *Makromol. Chem.* **193**, 1127–1133 (1992).

## Acknowledgements

Authors acknowledge A. Putintsev for technical support. This work was supported by the Russian Science Foundation (RSF) grant no. 20-72-10145 and the UK's Engineering and Physical Sciences Research Council grant EP/M025330/1 on Hybrid Polaritonics. E.S.A. and V.Yu.Sh. thank the Foundation for the Advancement of Theoretical Physics and Mathematics Basis. Yu.E.L. acknowledges Basic Research Program at the National Research University HSE, D.U., F.S. and T.S. acknowledge support by QuantERA project RouTe (SNSF grant no. 20QT21 175389). P.G.L, D.U., T.S. and R.F.M. acknowledge support by European H2020-FETOPEN project POLLOC (Grant No. 899141).

## Author information

### Affiliations

1. Center for Photonics and Quantum Materials, Skolkovo Institute of Science and Technology, Moscow, Russia

Anton V. Zasedatelev, Anton V. Baranikov, Denis Sannikov, Vladislav Yu. Shishkov, Evgeny S. Andrianov, Yurii E. Lozovik & Pavlos G. Lagoudakis

2. Laboratories for Hybrid Photonics, Skolkovo Institute of Science and Technology, Moscow, Russia

Anton V. Zasedatelev, Anton V. Baranikov, Denis Sannikov, Vladislav Yu. Shishkov, Evgeny S. Andrianov, Yurii E. Lozovik & Pavlos G. Lagoudakis

3. IBM Research Europe - Zurich, Rüschlikon, Switzerland

Darius Urbonas, Fabio Scafirimuto, Thilo Stöferle & Rainer F. Mahrt

4. Dukhov Research Institute of Automatics (VNIIA), Moscow, Russia

Vladislav Yu. Shishkov, Evgeny S. Andrianov & Yurii E. Lozovik

5. Moscow Institute of Physics and Technology, Dolgoprudny, Russia

Vladislav Yu. Shishkov & Evgeny S. Andrianov

6. Institute for Spectroscopy RAS, Troitsk, Russia

Yurii E. Lozovik

7. Moscow Institute of Electronics and Mathematics, National Research University Higher School of Economics, Moscow, Russia

Yurii E. Lozovik

8. Macromolecular Chemistry Group and Institute for Polymer Technology, Bergische Universität Wuppertal, Wuppertal, Germany

Ullrich Scherf

9. Department of Physics and Astronomy, University of Southampton, Southampton, UK

Pavlos G. Lagoudakis

## Contributions

A.V.Z., A.V.B. and D.S. performed the experiments and analysed the data. D.U., F.S., T.S. and R.F.M. contributed to the design and fabrication of the

organic microcavity. U.S. synthesized the organic material. V.Yu.Sh., E.S.A. and Yu.E.L. developed microscopic theory and carried out numerical simulations. A.V.Z. and P.G.L. designed and led the research. The manuscript was written through contributions from all authors. All authors have given approval to the final version of the manuscript.

## Corresponding authors

Correspondence to Anton V. Zasedatelev or Pavlos G. Lagoudakis.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

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

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

### Supplementary Information

Supplementary Sections 1–7, including text and data, Supplementary Figs. 1–22, Table 1 and references.

### Peer Review File

## Rights and permissions

## [Reprints and Permissions](#)

## About this article

### Cite this article

Zasedatelev, A.V., Baranikov, A.V., Sannikov, D. *et al.* Single-photon nonlinearity at room temperature. *Nature* **597**, 493–497 (2021).  
<https://doi.org/10.1038/s41586-021-03866-9>

- Received: 25 December 2020
- Accepted: 29 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03866-9>

### 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

## [Light detection nears its quantum limit](#)

- Sebastian Klembt

News & Views 22 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03866-9>

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

- Article
- [Published: 22 September 2021](#)

# Electron phase-space control in photonic chip-based particle acceleration

- [R. Shiloh](#) [ORCID: orcid.org/0000-0001-5142-7467](#)<sup>1 na1</sup>,
- [J. Illmer](#) [ORCID: orcid.org/0000-0003-2146-7366](#)<sup>1 na1</sup>,
- [T. Chlouba](#) [ORCID: orcid.org/0000-0002-8383-1293](#)<sup>1 na1</sup>,
- [P. Yousefi](#)<sup>1</sup>,
- [N. Schönenberger](#)<sup>1,2</sup>,
- [U. Niedermayer](#) [ORCID: orcid.org/0000-0002-7671-980X](#)<sup>3</sup>,
- [A. Mittelbach](#)<sup>1</sup> &
- [P. Hommelhoff](#) <sup>1,2</sup>

[Nature](#) volume **597**, pages 498–502 (2021)

- 929 Accesses
- 59 Altmetric
- [Metrics details](#)

## Subjects

- [Nanophotonics and plasmonics](#)
- [Photonic devices](#)

## Abstract

Particle accelerators are essential tools in science, hospitals and industry<sup>1,2,3,4,5,6</sup>. Yet their costs and large footprint, ranging in length from metres to several kilometres, limit their use. The recently demonstrated nanophotonics-based acceleration of charged particles can reduce the cost and size of these accelerators by orders of magnitude<sup>7,8,9</sup>. In this approach, a carefully designed nanostructure transfers energy from laser light to the particles in a phase-synchronous manner, accelerating them. To accelerate particles to the megaelectronvolt range and beyond, with minimal particle loss<sup>10,11</sup>, the particle beam needs to be confined over extended distances, but the necessary control of the electron beam's phase space has been elusive. Here we demonstrate complex electron phase-space control at optical frequencies in the 225-nanometre narrow channel of a silicon-based photonic nanostructure that is 77.7 micrometres long. In particular, we experimentally show alternating phase focusing<sup>10,11,12,13</sup>, a particle propagation scheme for minimal-loss transport that could, in principle, be arbitrarily long. We expect this work to enable megaelectronvolt electron-beam generation on a photonic chip, with potential for applications in radiotherapy and compact light sources<sup>9</sup>, and other forms of electron phase-space control resulting in narrow energy or zeptosecond-bunched beams<sup>14,15,16</sup>.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

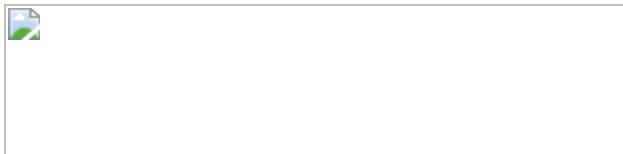
[Rent or Buy](#)

All prices are NET prices.

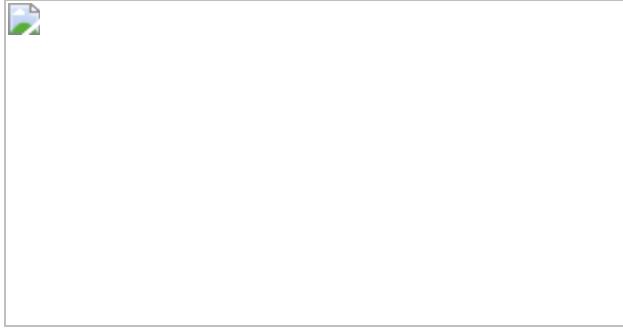
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Forces acting as a function of the synchronous phase  $\phi_s$ .**



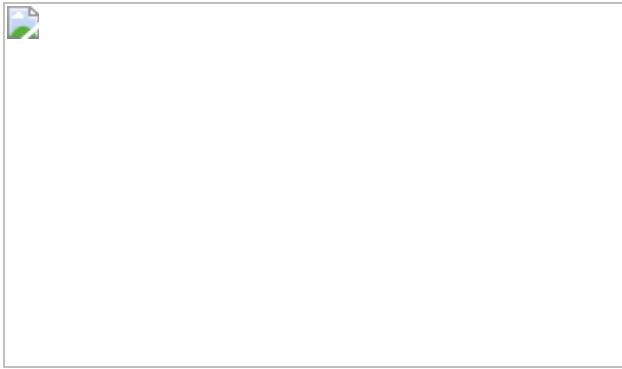
**Fig. 2: Complex optical electron phase-space control in alternating phase focusing.**



**Fig. 3: Silicon photonic nanostructure for phase-space control.**



**Fig. 4: Experimental verification of the APF scheme.**



## Data availability

Source data for Fig. 4a, c are provided with the paper. The data in Fig. 4b, d that support the findings of this study are available in Zenodo with the identifier <https://doi.org/10.5281/zenodo.4446542>. Source data are provided with this paper.

## References

1. 1.

Aad, G. et al. A particle consistent with the Higgs boson observed with the ATLAS detector at the large hadron collider. *Science* **338**, 1576–1582 (2012).

2. 2.

Chatrchyan, S. et al. A new boson with a mass of 125 GeV observed with the CMS experiment at the large hadron collider. *Science* **338**, 1569–1575 (2012).

3. 3.

Focus: Synchrotron techniques. *Nat. Rev. Mater.* <https://www.nature.com/collections/vjzmtcbvzy> (2018).

4. 4.

Bucksbaum, P., Möller, T. & Ueda, K. Frontiers of free-electron laser science. *J. Phys. At. Mol. Opt. Phys.* **46**, 160201 (2013).

5. 5.

Karzmark, C. J. Advances in linear accelerator design for radiotherapy. *Med. Phys.* **11**, 105–128 (1984).

6. 6.

Podgorsak, E. B. *Radiation Oncology Physics: A Handbook for Teachers and Students* (International Atomic Energy Agency, 2005).

7. 7.

Breuer, J. & Hommelhoff, P. Laser-based acceleration of nonrelativistic electrons at a dielectric structure. *Phys. Rev. Lett.* **111**, 134803 (2013).

8. 8.

Peralta, E. A. et al. Demonstration of electron acceleration in a laser-driven dielectric microstructure. *Nature* **503**, 91–94 (2013).

9. 9.

England, R. J. et al. Dielectric laser accelerators. *Rev. Mod. Phys.* **86**, 1337–1389 (2014).

10. 10.

Wangler, T. P. *RF Linear Accelerators* 2nd edition (Wiley-VCH, 2008).

11. 11.

Chao, A. W., Mess, K. H., Tigner, M. & Zimmermann, F. *Handbook of Accelerator Physics and Engineering* 2nd edition (2013).

12. 12.

Courant, E. D. & Snyder, H. S. Theory of the alternating-gradient synchrotron. *Ann. Phys.* **3**, 1–48 (1958).

13. 13.

Niedermayer, U., Egenolf, T., Boine-Frankenheim, O. & Hommelhoff, P. Alternating-phase focusing for dielectric-laser acceleration. *Phys. Rev. Lett.* **121**, 214801 (2018).

14. 14.

Schönenberger, N. et al. Generation and characterization of attosecond microbunched electron pulse trains via dielectric laser acceleration. *Phys. Rev. Lett.* **123**, 264803 (2019).

15. 15.

Black, D. S. et al. Net acceleration and direct measurement of attosecond electron pulses in a silicon dielectric laser accelerator. *Phys. Rev. Lett.* **123**, 264802 (2019).

16. 16.

Niedermayer, U. et al. Low-energy-spread attosecond bunching and coherent electron acceleration in dielectric nanostructures. *Phys. Rev. Appl.* **15**, L021002 (2021).

17. 17.

Panofsky, W. K. H. & Wenzel, W. A. Some considerations concerning the transverse deflection of charged particles in radio-frequency fields. *Rev. Sci. Instrum.* **27**, 967 (1956).

18. 18.

Shimoda, K. Proposal for an electron accelerator using an optical maser. *Appl. Opt.* **1**, 33–35 (1962).

19. 19.

Lohmann, A. Electron acceleration by light waves. *IBM Tech. Note* **5**, 169–182 (1962).

20. 20.

Cesar, D. et al. High-field nonlinear optical response and phase control in a dielectric laser accelerator. *Commun. Phys.* **1**, 46 (2018).

21. 21.

Leedle, K. J., Fabian Pease, R., Byer, R. L. & Harris, J. S. Laser acceleration and deflection of 96.3 keV electrons with a silicon dielectric structure. *Optica* **2**, 158–161 (2015).

22. 22.

Kozák, M. et al. Optical gating and streaking of free electrons with sub-optical cycle precision. *Nat. Commun.* **8**, 14342 (2017).

23. 23.

Leedle, K. J. et al. Phase-dependent laser acceleration of electrons with symmetrically driven silicon dual pillar gratings. *Opt. Lett.* **43**, 2181 (2018).

24. 24.

McNeur, J. et al. Elements of a dielectric laser accelerator. *Optica* **5**, 687–690 (2018).

25. 25.

Black, D. S. et al. Laser-driven electron lensing in silicon microstructures. *Phys. Rev. Lett.* **122**, 104801 (2019).

26. 26.

Sapra, N. V. et al. On-chip integrated laser-driven particle accelerator. *Science* **367**, 79–83 (2020).

27. 27.

Shiltsev, V. & Zimmermann, F. Modern and future colliders. *Rev. Mod. Phys.* **93**, 015006 (2021).

28. 28.

Naranjo, B., Valloni, A., Puterman, S. & Rosenzweig, J. B. Stable charged-particle acceleration and focusing in a laser accelerator using spatial harmonics. *Phys. Rev. Lett.* **109**, 164803 (2012).

29. 29.

Niedermayer, U., Egenolf, T. & Boine-Frankenheim, O. Three dimensional alternating-phase focusing for dielectric-laser electron accelerators. *Phys. Rev. Lett.* **125**, 164801 (2020).

30. 30.

Kozák, M. et al. Ultrafast scanning electron microscope applied for studying the interaction between free electrons and optical near-fields of periodic nanostructures. *J. Appl. Phys.* **124**, 023104 (2018).

31. 31.

Yousefi, P. et al. Dielectric laser electron acceleration in a dual pillar grating with a distributed Bragg reflector. *Opt. Lett.* **44**, 1520 (2019).

32. 32.

Roques-Carmes, C. et al. Towards integrated tunable all-silicon free-electron light sources. *Nat. Commun.* **10**, 3176 (2019).

33. 33.

Wiedemann, H. *Particle Accelerator Physics* (Springer, 2015).

34. 34.

Hirano, T. et al. A compact electron source for the dielectric laser accelerator. *Appl. Phys. Lett.* (2020).

35. 35.

Zhao, Z. et al. Design of a multichannel photonic crystal dielectric laser accelerator. *Photon. Res.* **8**, 1586–1598 (2020).

36. 36.

Staude, I. et al. Waveguides in three-dimensional photonic bandgap materials for particle-accelerator on a chip architectures. *Opt. Express* **20**, 5607–5612 (2012).

37. 37.

Egenolf, T., Niedermayer, U. & Boine-Frankenheim, O. Tracking with wakefields in dielectric laser acceleration grating structures. *Phys. Rev. Accel. Beams* **23**, 054402 (2020). 0987654321``

38. 38.

Kimura, W. D., Poaorelsky, I. V & Schächter, L. CO<sub>2</sub>-laser-driven dielectric laser accelerator. In *2018 IEEE Advanced Accelerator Concepts Workshop (AAC)* <https://doi.org/10.1109/AAC.2018.8659403> (2018).

39. 39.

Egerton, R. F. Outrun radiation damage with electrons? *Adv. Struct. Chem. Imaging* **1**, 5 (2015).

40. 40.

Cros, B. & Muggli, P. Input to the European Particle Physics Strategy Update. in *Advanced Linear Collider Study Group (ALEGRO collaboration)* (2018).

41. 41.

Brüning, O. S. et al. *LHC Design Report* CERN-2004-003-V-1 (2004).

## Acknowledgements

We acknowledge discussions with the members of the Accelerator on a Chip International Program (ACCHIP). We thank the clean-room facility staff at the Max Planck Institute for the Science of Light for continued assistance. We acknowledge funding by the Gordon and Betty Moore Foundation (#GBMF4744), ERC grants NearFieldAtto (#616823) and AccelOnChip (#884217) and BMBF projects 05K19WEB and 05K19RDE.

## Author information

### Author notes

1. These authors contributed equally: R. Shiloh, J. Illmer, T. Chlouba

## Affiliations

1. Physics Department, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

R. Shiloh, J. Illmer, T. Chlouba, P. Yousefi, N. Schönenberger, A. Mittelbach & P. Hommelhoff

2. Max Planck Institute for the Science of Light, Erlangen, Germany

N. Schönenberger & P. Hommelhoff

3. Technische Universität Darmstadt, Institute for Accelerator Science and Electromagnetic Fields (TEMF), Darmstadt, Germany

U. Niedermayer

## Contributions

T.C. and J.I. measured the data. R.S. and U.N. designed the structures and performed simulations. P.Y. fabricated the structures. J.I., R.S. and T.C. analysed the data. N.S. and A.M. inferred stringent tolerance requirements from initial measurements. J.I., R.S. and P.H. wrote the manuscript. P.H. supervised the experiment.

## Corresponding authors

Correspondence to R. Shiloh or J. Illmer or P. Hommelhoff.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks James Rosenzweig, Yelong Wei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Photonic nanostructure on top of a mesa: over-focusing structure.](#)

This SEM image shows a structure fabricated on top of a 60  $\mu\text{m}$ -high mesa. The photonic structure for the over-focusing measurement is visible atop the mesa. The five gaps between the macro cells are also clearly visible. Here, 24 pillar pairs build one macro cell. Input (left) and output (right) apertures (thick blocks) are used for alignment during experiment.

Additional pillars to the left of the input aperture act as markers that identify the specific structure during the experiment. The mesa allows us to focus the laser beam under  $0^\circ$  incidence angle from the side (see also Extended Data Fig. [1c](#)).

### **Extended Data Fig. 2 Schematic of the experimental set-up.**

**(a)** Both IR (red) and UV (blue) laser pulses are generated with the help of an optical parametric amplifier (OPA). The UV pulses are focused onto the Schottky emitter of the SEM, where they release electron pulses. The electron pulses pass through the electron column and are focused into the nano-photonic channel. The IR pulses first pass a neutral-density (ND) filter for variable optical power attenuation before they traverse a delay stage, where the time delay between IR laser pulses and electron pulses is set. A bandpass filter limits the IR spectrum so that the laser pulses are stretched to a duration of 680 fs (FWHM). A cylindrical telescope is used to generate an elliptic laser beam with a 1:6 ratio. The beam is then split into two parts, where one part is used to monitor the power during measurements and the other is focused on the structure via an aspherical lens (ASL). The back reflection from the sample is used to align the laser beam to the structure. The electron energy is measured with a magnetic deflection spectrometer and an MCP detector with a phosphor screen, viewed from outside the vacuum chamber (indicated by the dashed line) with a CCD camera (not shown). For structures without a mesa **(b)** the incident angle of the laser beam is  $5^\circ$ , while for structures on top of the mesa **(c)** this angle is  $0^\circ$ .

### **Extended Data Fig. 3 Photonic nanostructure on flat substrate: high contrast structure.**

The scanning electron microscope (SEM) image shows the dual pillar transport structure for high contrast measurements. The dual pillar transport channel can be seen on the right as a colonnade structure. The four solid slabs left-above of the colonnades structure are a distributed Bragg mirror<sup>[31](#)</sup>. An alignment aperture is placed at the input of the structure (thick blocks). Electrons are focused into the colonnade structure, that is, into the channel between the rows of pillars. The laser beam impinges on the

structure from the side, from bottom-right here, perpendicular to the pillars and with a 5° angle to the substrate (see also Extended Data Fig. [1b](#)).

## Extended Data Fig. 4 Particle tracking simulation: One optical phase vs. all optical phases.

(a) Example particle trajectories for the optimal guiding field strength. The differently colored curves in between the pillars denote individual particle trajectories. The APF behaviour is clearly visible in the breathing of the envelope of the particle trajectories. For illustrative purposes the simulation was conducted with an electron pulse length of 0.001 fs. This way, the APF effect is apparent because electrons only sample fields of a small random fraction of the optical cycle, away from the crest. The lattice periodicity of  $2 \times (7.149 \mu\text{m} + 0.589 \mu\text{m}) = 15.47 \mu\text{m}$  (see Fig. [3](#)), which is equal to the beta function periodicity, can be directly seen in the envelope of the trajectories, where the phase jump is manifested as a bending of the curve. Hence we can here directly assign F to macro cell 2, D to macro cell 3, and so on repeatedly (see numbering at the top of each macro cell, macro cell 1 consists of 2 half cells at the beginning and the end of the structure). (b) Same as (a) but with all optical phases uniformly sampled. The electron pulse (flat-hat here) is one optical period long. Evidently, the existence of the two fixed points in phase,  $\pi/2$  and  $3\pi/2$ , cannot be easily discerned close to the entrance of the structure, meaning that F and D cannot be uniquely assigned to each cell. Halfway through the structure, however, the lattice oscillations become visible again in the form of two “modes”, namely two trajectory classes, each directly linked to the lattice period again. The evolution into this two-mode structure is a consequence of the electrons accumulating around the two fixed points in phase space separated by  $\pi^{[13](#)}$ . The result is an overlapping of two breathing motions in the trajectories shifted by one macro cell. (c) Same as (b) though the flat-hat electron pulse is here 400 fs long, the APF scheme still leads to particle propagation with hardly any loss, and the overall envelope nearly matches that of the single-cycle pulse in (b) (same as Fig. [4e](#) of the main text). (d) Example particle trajectories with the over-focusing behaviour visible (same as Fig. [4f](#) of the main text). Again, we only show particle trajectories of an electron pulse with a duration of 0.001 fs. Particles are obviously lost, where the lost particles’ deviation from the design axis exceeds the aperture of the

structure. (e) Same as in (d) but with all phases sampled. A similar effect as depicted in (b) takes place, where the loss of particles at the structure boundary occurs at multiple locations. The reason is, again, that macro cells act differently depending on where the electron is with respect to the optical phase, and hence experience forces shifted in phase. Most importantly, the overall performance of the APF scheme is maintained, if not for the limitation of the beam size due to the structure aperture. The colours in panels a, c, and d were chosen so individual trajectories can be better discerned. The colours in b and e indicate the injection time of the electron (see colour bar).

## Source data

### [Source Data Fig. 4](#)

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Shiloh, R., Illmer, J., Chlouba, T. *et al.* Electron phase-space control in photonic chip-based particle acceleration. *Nature* **597**, 498–502 (2021). <https://doi.org/10.1038/s41586-021-03812-9>

- Received: 29 January 2021
- Accepted: 07 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03812-9>

## 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/s41586-021-03812-9>

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

- Article
- [Published: 22 September 2021](#)

# Three-dimensional electronic microfliers inspired by wind-dispersed seeds

- [Bong Hoon Kim](#)<sup>1,2</sup> na1,
- [Kan Li](#)<sup>3,4</sup> na1,
- [Jin-Tae Kim](#) [ORCID: orcid.org/0000-0001-9933-1931](#)<sup>5</sup> na1,
- [Yoonseok Park](#) [ORCID: orcid.org/0000-0002-1702-0986](#)<sup>5</sup> na1,
- [Hokyung Jang](#) [ORCID: orcid.org/0000-0002-7797-9881](#)<sup>6</sup>,
- [Xueju Wang](#)<sup>7</sup>,
- [Zhaoqian Xie](#) [ORCID: orcid.org/0000-0003-1320-817X](#)<sup>8,9</sup>,
- [Sang Min Won](#) [ORCID: orcid.org/0000-0002-5750-8628](#)<sup>10</sup>,
- [Hong-Joon Yoon](#)<sup>5</sup>,
- [Geumbee Lee](#)<sup>5</sup>,
- [Woo Jin Jang](#)<sup>11</sup>,
- [Kun Hyuck Lee](#)<sup>5</sup>,
- [Ted S. Chung](#) [ORCID: orcid.org/0000-0002-0212-1071](#)<sup>5</sup>,
- [Yei Hwan Jung](#)<sup>12</sup>,
- [Seung Yun Heo](#)<sup>5</sup>,
- [Yechan Lee](#) [ORCID: orcid.org/0000-0002-4744-267X](#)<sup>13</sup>,
- [Juyun Kim](#)<sup>11</sup>,
- [Tengfei Cai](#)<sup>14</sup>,
- [Yeonha Kim](#)<sup>11</sup>,
- [Poom Prasopsukh](#)<sup>14</sup>,
- [Yongjoon Yu](#) [ORCID: orcid.org/0000-0003-0142-1600](#)<sup>5</sup>,
- [Xinge Yu](#) [ORCID: orcid.org/0000-0003-0522-1171](#)<sup>15</sup>,
- [Raudel Avila](#)<sup>16,17,18</sup>,

- [Haiwen Luan](#)<sup>5,16,17,18</sup>,
- [Honglie Song](#)<sup>19</sup>,
- [Feng Zhu](#)<sup>20</sup>,
- [Ying Zhao](#)<sup>21</sup>,
- [Lin Chen](#)<sup>22</sup>,
- [Seung Ho Han](#)<sup>23</sup>,
- [Jiwoong Kim](#)<sup>1,2</sup>,
- [Soong Ju Oh](#)<sup>24</sup>,
- [Heon Lee](#) ORCID: [orcid.org/0000-0003-1964-3746](https://orcid.org/0000-0003-1964-3746)<sup>24</sup>,
- [Chi Hwan Lee](#) ORCID: [orcid.org/0000-0002-4868-7054](https://orcid.org/0000-0002-4868-7054)<sup>25,26,27</sup>,
- [Yonggang Huang](#) ORCID: [orcid.org/0000-0002-0483-8359](https://orcid.org/0000-0002-0483-8359)<sup>16,17,18</sup>,
- [Leonardo P. Chamorro](#) ORCID: [orcid.org/0000-0002-5199-424X](https://orcid.org/0000-0002-5199-424X)<sup>14</sup>,
- [Yihui Zhang](#) ORCID: [orcid.org/0000-0003-0885-2067](https://orcid.org/0000-0003-0885-2067)<sup>19</sup> &
- [John A. Rogers](#) ORCID: [orcid.org/0000-0002-2980-3961](https://orcid.org/0000-0002-2980-3961)<sup>5,17,18,28,29,30,31</sup>

*Nature* volume **597**, pages 503–510 (2021)

- 4725 Accesses
- 893 Altmetric
- [Metrics details](#)

## Subjects

- [Aerospace engineering](#)
- [Electronic devices](#)
- [Fluid dynamics](#)
- [Fluidics](#)
- [Mechanical engineering](#)

## Abstract

Large, distributed collections of miniaturized, wireless electronic devices<sup>1,2</sup> may form the basis of future systems for environmental monitoring<sup>3</sup>,

population surveillance<sup>4</sup>, disease management<sup>5</sup> and other applications that demand coverage over expansive spatial scales. Aerial schemes to distribute the components for such networks are required, and—inspired by wind-dispersed seeds<sup>6</sup>—we examined passive structures designed for controlled, unpowered flight across natural environments or city settings. Techniques in mechanically guided assembly of three-dimensional (3D) mesostructures<sup>7,8,9</sup> provide access to miniature, 3D fliers optimized for such purposes, in processes that align with the most sophisticated production techniques for electronic, optoelectronic, microfluidic and microelectromechanical technologies. Here we demonstrate a range of 3D macro-, meso- and microscale fliers produced in this manner, including those that incorporate active electronic and colorimetric payloads. Analytical, computational and experimental studies of the aerodynamics of high-performance structures of this type establish a set of fundamental considerations in bio-inspired design, with a focus on 3D fliers that exhibit controlled rotational kinematics and low terminal velocities. An approach that represents these complex 3D structures as discrete numbers of blades captures the essential physics in simple, analytical scaling forms, validated by computational and experimental results. Battery-free, wireless devices and colorimetric sensors for environmental measurements provide simple examples of a wide spectrum of applications of these unusual concepts.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

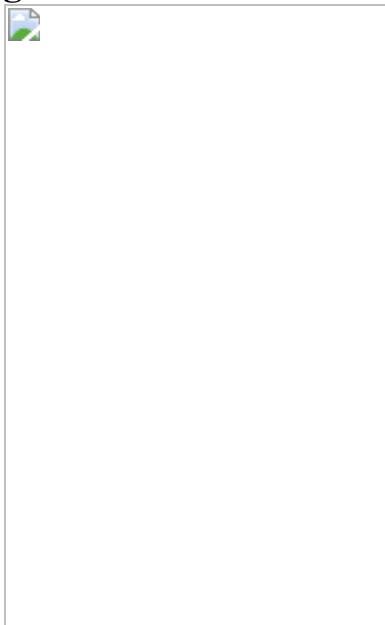
[Rent or Buy](#)

All prices are NET prices.

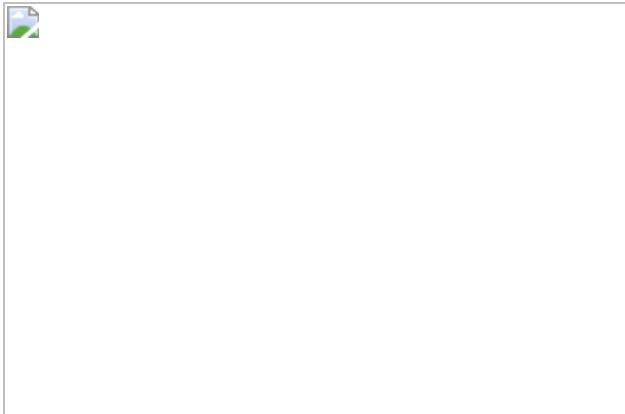
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

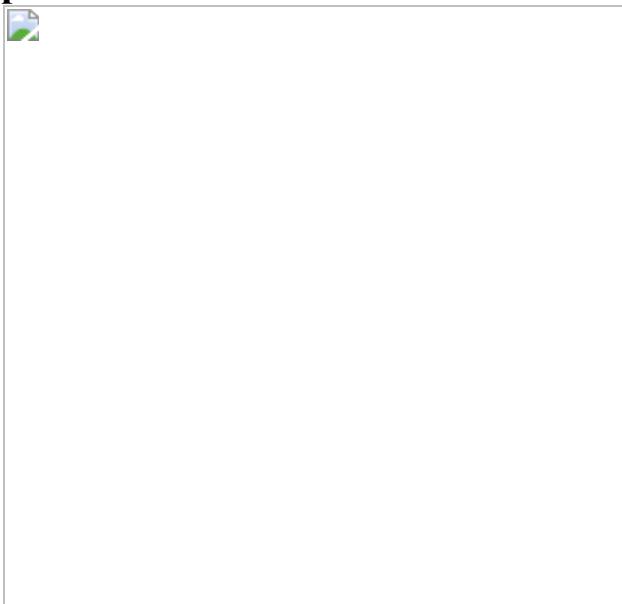
**Fig. 1: 3D microfliers inspired by wind-dispersed seeds.**



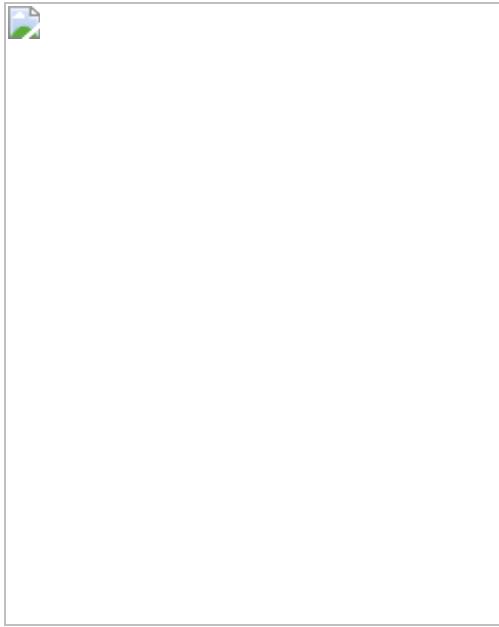
**Fig. 2: Theoretical analysis and numerical simulation of the aerodynamics associated with representative 3D micro-, meso- and macrofliers.**



**Fig. 3: Experimental measurements of the flow characteristics of representative 3D mesofliers.**



**Fig. 4: 3D colorimetric mesofliers, electronic mesofliers and IoT macrofliers.**



## Data availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

## References

1. 1. Chung, H. U. et al. Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care. *Science* **363**, eaau0780 (2019).
2. 2. Kim, B. H. et al. Mechanically guided post-assembly of 3D electronic systems. *Adv. Funct. Mater.* **28**, 1803149 (2018).
3. 3. Jin, J., Wang, Y., Jiang, H. & Chen, X. Evaluation of microclimatic detection by a wireless sensor network in forest ecosystems. *Sci. Rep.*

**8**, 16433 (2018).

4. 4.

González-Alcaide, G., Llorente, P. & Ramos-Rincón, J. M. Systematic analysis of the scientific literature on population surveillance. *Helyon* **6**, E05141 (2020).

5. 5.

Groseclose, S. L. & Buckeridge, D. L. Public health surveillance systems: recent advances in their use and evaluation. *Annu. Rev. Public Health* **38**, 57–79 (2017).

6. 6.

Augspurger, C. K. Morphology and dispersal potential of wind-dispersed diaspores of neotropical trees. *Am. J. Bot.* **73**, 353–363 (1986).

7. 7.

Won, S. M. et al. Multimodal sensing with a three-dimensional piezoresistive structure. *ACS Nano* **13**, 10972–10979 (2019).

8. 8.

Kim, B. H. et al. Three-dimensional silicon electronic systems fabricated by compressive buckling process. *ACS Nano* **12**, 4164–4171 (2018).

9. 9.

Park, Y. et al. Three-dimensional, multifunctional neural interfaces for cortical spheroids and engineered assembloids. *Sci. Adv.* **7**, eabf9153 (2021).

10. 10.

Nathan, R. et al. Mechanisms of long-distance dispersal of seeds by wind. *Nature* **418**, 409–413 (2002).

11. 11.

Nathan, R. Long-distance dispersal of plants. *Science* **313**, 786–788 (2006).

12. 12.

Seale, M. & Nakayama, N. From passive to informed: mechanical mechanisms of seed dispersal. *New Phytol.* **225**, 653–658 (2020).

13. 13.

Cummins, C. et al. A separated vortex ring underlies the flight of the dandelion. *Nature* **562**, 414–418 (2018).

14. 14.

Rabault, J., Fauli, R. A. & Carlson, A. Curving to fly: synthetic adaptation unveils optimal flight performance of whirling fruits. *Phys. Rev. Lett.* **122**, 024501 (2019).

15. 15.

Fauli, R. A., Rabault, J. & Carlson, A. Effect of wing fold angles on the terminal descent velocity of double-winged autorotating seeds, fruits, and other diaspores. *Phys. Rev. E* **100**, 013108 (2019).

16. 16.

Xu, S. et al. Assembly of micro/nanomaterials into complex, three-dimensional architectures by compressive buckling. *Science* **347**, 154–159 (2015).

17. 17.

Zhang, Y. et al. A mechanically driven form of kirigami as a route to 3D mesostructures in micro/nanomembranes. *Proc. Natl Acad. Sci. USA* **112**, 11757–11764 (2015).

18. 18.

Zhang, Y. et al. Printing, folding and assembly methods for forming 3D mesostructures in advanced materials. *Nat. Rev. Mater.* **2**, 17019 (2017).

19. 19.

Pang, W. et al. Electro-mechanically controlled assembly of reconfigurable 3D mesostructures and electronic devices based on dielectric elastomer platforms. *Natl Sci. Rev.* **7**, 342–354 (2020).

20. 20.

Wang, X. et al. Freestanding 3D mesostructures, functional devices, and shape-programmable systems based on mechanically induced assembly with shape memory polymers. *Adv. Mater.* **31**, 1805615 (2019).

21. 21.

Greene, D. F. & Johnson, E. A. Seed mass and dispersal capacity in wind-dispersed diaspores. *Oikos* **67**, 69–74 (1993).

22. 22.

Zhang, L., Wang, X., Moran, M. D. & Feng, J. Review and uncertainty assessment of size-resolved scavenging coefficient formulations for below-cloud snow scavenging of atmospheric aerosols. *Atmos. Chem. Phys.* **13**, 10005–10025 (2013).

23. 23.

Hölzer, A. & Sommerfeld, M. New simple correlation formula for the drag coefficient of non-spherical particles. *Powder Technol.* **184**, 361–

365 (2008).

24. 24.

Belmonte, A., Eisenberg, H. & Moses, E. From flutter to tumble: inertial drag and Froude similarity in falling paper. *Phys. Rev. Lett.* **81**, 345–348 (1998).

25. 25.

Zhong, H. et al. Experimental investigation of freely falling thin disks. Part 1. The flow structures and Reynolds number effects on the zigzag motion. *J. Fluid Mech.* **716**, 228–250 (2013).

26. 26.

Kim, J. T., Jin, Y., Shen, S., Dash, A. & Chamorro, L. P. Free fall of homogeneous and heterogeneous cones. *Phys. Rev. Fluids* **5**, 093801 (2020).

27. 27.

Chigurupati, N. et al. Evaluation of red cabbage dye as a potential natural color for pharmaceutical use. *Int. J. Pharm.* **241**, 293–299 (2002).

28. 28.

Binnig, J., Meyer, J. & Kasper, G. Calibration of an optical particle counter to provide PM2.5 mass for well-defined particle materials. *J. Aerosol Sci.* **38**, 325–332 (2007).

29. 29.

Shao, W., Zhang, H. & Zhou, H. Fine particle sensor based on multi-angle light scattering and data fusion. *Sensors* **17**, 1033 (2017).

30. 30.

Heo, S. Y. et al. Wireless, battery-free, flexible, miniaturized dosimeters monitor exposure to solar radiation and to light for phototherapy. *Sci. Transl. Med.* **10**, eaau1643 (2018).

31. 31.

Kang, S. K., Koo, J., Lee, Y. K. & Rogers, J. A. Advanced materials and devices for bioresorbable electronics. *Acc. Chem. Res.* **51**, 988–998 (2018).

32. 32.

Lee, G., Choi, Y. S., Yoon, H. J. & Rogers, J. A. Advances in physicochemically stimuli-responsive materials for on-demand transient electronic systems. *Matter* **3**, 1031–1052 (2020).

33. 33.

Kim, J. T., Nam, J., Shen, S., Lee, C. & Chamorro, L. P. On the dynamics of air bubbles in Rayleigh-Bénard convection. *J. Fluid Mech.* **891**, A7 (2020).

34. 34.

Kim, J. T. & Chamorro, L. P. Lagrangian description of the unsteady flow induced by a single pulse of a jellyfish. *Phys. Rev. Fluids* **4**, 064605 (2019).

## Acknowledgements

This work was supported by the Querrey Simpson Institute for Bioelectronics at Northwestern University. B.H.K. acknowledges support from the following: National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (nos 2019R1G1A1100737, 2020R1C1C1014980); the Nanomaterial Technology Development Program (NRF-2016M3A7B4905613) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning; the Project for Collabo R&D between Industry, Academy

and Research Institute funded by Korean Ministry of SMEs and Startups in 2020/2021 (project no. S2890749/S3104531); the Nano Material Technology Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2009-0082580); the National Research Facilities and Equipment Center at the Ministry of Science and ICT (Support Program for Equipment Transfer, grant no. 1711116699); and the Glint Materials Company. K.L. acknowledges support from the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology (grant no. DMETKF2021010). Y.P. acknowledges support from the German Research Foundation (PA 3154/1-1). Y.Z. acknowledges support from the National Natural Science Foundation of China (grant no. 12050004), the Institute for Guo Qiang, Tsinghua University (grant no. 2019GQG1012), and the Tsinghua National Laboratory for Information Science and Technology. H.L. acknowledges support from the Creative Materials Discovery Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2018M3D1A1058972). S.M.W. acknowledges support from the National Research Foundation of Korea funded by the Ministry of Science and ICT of Korea (NRF-2021M3H4A1A01079367), and by the Nano Material Technology Development Program (2020M3H4A1A03084600) funded by the Ministry of Science and ICT of Korea. Y.H.J. acknowledges support from the research fund of Hanyang University (HY-20210000000832). C.H.L. acknowledges funding support from the National Science Foundation (2032529-CBET). Z.X. acknowledges support from the National Natural Science Foundation of China (grant no. 12072057), the LiaoNing Revitalization Talents Program (grant no. XLYC2007196), and Fundamental Research Funds for the Central Universities (grant no. DUT20RC(3)032). R.A. acknowledges support from the National Science Foundation Graduate Research Fellowship (NSF grant number 1842165) and a Ford Foundation Predoctoral Fellowship. We thank Jaeeun Koo for artwork in Fig. [1a](#).

## Author information

### Author notes

1. These authors contributed equally: Bong Hoon Kim, Kan Li, Jin-Tae Kim, Yoonseok Park

## Affiliations

1. Department of Organic Materials and Fiber Engineering, Soongsil University, Seoul, Republic of Korea  
Bong Hoon Kim & Jiwoong Kim
2. Department of Smart Wearable Engineering, Soongsil University, Seoul, Republic of Korea  
Bong Hoon Kim & Jiwoong Kim
3. Department of Engineering, University of Cambridge, Cambridge, UK  
Kan Li
4. State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, People's Republic of China  
Kan Li
5. Querrey Simpson Institute for Bioelectronics, Northwestern University, Evanston, IL, USA  
Jin-Tae Kim, Yoonseok Park, Hong-Joon Yoon, Geumbee Lee, Kun Hyuck Lee, Ted S. Chung, Seung Yun Heo, Yongjoon Yu, Haiwen Luan & John A. Rogers
6. Department of Electrical and Computer Engineering, University of Wisconsin Madison, Madison, WI, USA  
Hokyung Jang
7. Department of Materials Science and Engineering, Institute of Materials Science, University of Connecticut, Storrs, CT, USA

Xueju Wang

8. State Key Laboratory of Structural Analysis for Industrial Equipment, Department of Engineering Mechanics, Dalian University of Technology, Dalian, People's Republic of China

Zhaoqian Xie

9. Ningbo Institute of Dalian University of Technology, Ningbo, People's Republic of China

Zhaoqian Xie

10. Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Sang Min Won

11. Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL, USA

Woo Jin Jang, Juyun Kim & Yeonha Kim

12. Department of Electronic Engineering, Hanyang University, Seoul, Republic of Korea

Yei Hwan Jung

13. Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Yechan Lee

14. Department of Mechanical Science and Engineering, University of Illinois, Urbana, IL, USA

Tengfei Cai, Poom Prasopsukh & Leonardo P. Chamorro

15. Department of Biomedical Engineering, City University of Hong Kong, Hong Kong, China

Xinge Yu

16. Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA

Raudel Avila, Haiwen Luan & Yonggang Huang

17. Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

Raudel Avila, Haiwen Luan, Yonggang Huang & John A. Rogers

18. Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA

Raudel Avila, Haiwen Luan, Yonggang Huang & John A. Rogers

19. Applied Mechanics Laboratory, Department of Engineering Mechanics, Center for Flexible Electronics Technology, Tsinghua University, Beijing, People's Republic of China

Honglie Song & Yihui Zhang

20. School of Logistics Engineering, Wuhan University of Technology, Wuhan, People's Republic of China

Feng Zhu

21. School of Aerospace Engineering and Applied Mechanics, Tongji University, Shanghai, People's Republic of China

Ying Zhao

22. State Key Laboratory for Mechanical Behavior of Materials, School of Material Science and Engineering, Xi'an Jiaotong University, Xi'an, People's Republic of China

Lin Chen

23. Electronic Convergence Materials and Device Research Center, Korea Electronics Technology Institute, Seongnam, Republic of Korea

Seung Ho Han

24. Department of Materials Science and Engineering, Korea University, Seoul, Republic of Korea

Soong Ju Oh & Heon Lee

25. Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA

Chi Hwan Lee

26. School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA

Chi Hwan Lee

27. School of Materials Engineering, Purdue University, West Lafayette, IN, USA

Chi Hwan Lee

28. Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

John A. Rogers

29. Department of Neurological Surgery, Northwestern University, Evanston, IL, USA

John A. Rogers

30. Department of Chemistry, Northwestern University, Evanston, IL, USA

John A. Rogers

31. Department of Electrical Engineering and Computer Science,  
Northwestern University, Evanston, IL, USA

John A. Rogers

## Contributions

B.H.K., K.L., J.-T.K. and Y.P. contributed equally to this work. Y.H., L.P.C., Y.Z. and J.A.R. conceived the ideas and supervised the project. B.H.K., K.L., J.-T.K., Y.P., Y.H., L.P.C., Y.Z. and J.A.R. wrote the manuscript. H.J., X.W., S.M.W., H.-J.Y., G.L., W.J.J., K.H.L., Y.H.J., S.Y.H., Y.L., J.K., Y.K., Y.Y. and X.Y. performed microelectromechanical experiments. J.-T.K., T.C. and P.P. performed fluid dynamics experiments. Z.X., T.S.C., R.A., H.L., H.S., F.Z., Y.Z. and L.C. performed mechanical and electromagnetic modelling and simulation. S.H.H., J.K., S.J.O., H.L. and C.H.L. provided scientific and experimental advice. All authors commented on the manuscript.

## Corresponding authors

Correspondence to Yonggang Huang or Leonardo P. Chamorro or Yihui Zhang or John A. Rogers.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Elizabeth Helbling and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Bio-inspired 3D micro- and mesofliers.

Photographs of a  $10 \times 10$  array of 3D micro- and mesofliers.

### Extended Data Fig. 2 Schematic diagram of the configuration for computational fluid dynamics simulation.

(a) 3D rotational falling fliers and (b) 2D aerofoil.

### Extended Data Fig. 3 Comparison of $G_0$ and $G_1$ across all classes of fliers.

CFD simulation results for the components  $G_0$  and  $G_1$  of the drag coefficient  $C_D$ , for fliers of types R, H, M and PM, with the 2D disk as a comparison.

### Extended Data Fig. 4 3D microflier with porous design.

(a) Inspiration of porosity from nature: optical images of dandelion seeds and a feather. (b) FE simulated configuration of a 3D void-free microflier ( $p = 0$ ) and a 3D microflier of porosity design ( $p = 0.26$ ). (c) Images of scanned thickness of a 2D precursor for a porous microflier, with top view and perspective view, respectively. (d)  $G_{0(b)}$  and (e)  $G_{1(b)}$  versus the attack angle for various porosities. Normalized (f)  $G_{0(b)}$  and (g)  $G_{1(b)}$  over their void-free values versus porosity, with the CFD values of various  $\alpha$  in  $[0^\circ, 90^\circ]$  and analytic fittings.

### Extended Data Fig. 5 Mechanical simulation of a 3D microflier [3, H, 0.75].

Schematic images of (a) a parachute design where the blades have no rotational tilting, and (b) a rotating flier design with rotationally tilted blades. (c) Comparison of  $G_0$  and  $G_1$  between the parachute mode and rotational falling mode.

### Extended Data Fig. 6 Experimental setups.

(a) Schematic (top) and photograph (bottom) of 3D-PTV experiment on free-falling mesofliers and (b) schematic (top) and photograph (bottom) of high speed PIV experiment on fixed 3D IoT fliers above a wind tunnel.

### Extended Data Fig. 7 Changes in colour of a pH-responsive 3D mesoflier.

(a) Photographs of pH-responsive 3D mesoflier immersed in different buffer solutions with pH ranging from 3 to 11. (b) Response time of pH indicators after immersion into buffer solutions at different pH values.

### Extended Data Fig. 8 The electrical characteristics of Si NM n-channel transistor (channel width/length = 80/20 $\mu\text{m}$ ) and diode integrated with 3D mesofliers.

(a) Drain current as a function of source/drain voltage for the gate voltages from 0 to 3 V. (b) The log scale transfer curves as a function of gate voltage from -7 to 7 V. (c) Current-voltage characteristics of a diode.

### Extended Data Fig. 9 Experiments for particulate matter (PM).

(a) A dust generation chamber operated with kitchen blenders. Scanning electron microscope (SEM) images of fine dust generated by (b) corn starches, (c) incenses and (d) smoke cakes.

### Extended Data Fig. 10 Electromagnetic performance of coils for wireless power transmission.

**(a)** Normalized magnetic field generated by the commercial transmission antenna with dimensions (318 mm × 338 mm × 30 mm). **(b)** Magnetic field strength along the line (0,0,Z) as a function of the distance Z normal to the transmission antenna for different input power  $P_{\text{in}}$  (1, 4, and 8 W). **(c)** Scattering parameters for the electromagnetic energy transfer between the coils when the NFC coil is located at the centre of the transmission antenna (0,0,0). **(d)** Simulated power in the NFC coil at different distance Z normal to the primary antenna with  $P_{\text{in}} = 8$  W.

## Supplementary information

### Supplementary Information

This file includes 4 Supplementary Notes, 31 Supplementary Figures and 3 Supplementary Tables.

### Supplementary Video 1

Free fall of a *Tristellateia* seed.

### Supplementary Video 2

PIV experiment on a macroflier.

### Supplementary Video 3

A 2D precursor and a 3D mesoflier above the vertical wind tunnel.

### Supplementary Video 4

Free fall of a 2D precursor and a 3D microflier.

### Supplementary Video 5

Free fall of micro-, meso- and macrofliers.

## Supplementary Video 6

3D flow fields induced by a 2D precursor and a 3D mesoflier.

## Supplementary Video 7

Instantaneous flow fields induced by a 3D electronic flier.

## Peer Review File

## **Rights and permissions**

### Reprints and Permissions

## **About this article**

### **Cite this article**

Kim, B.H., Li, K., Kim, JT. *et al.* Three-dimensional electronic microfliers inspired by wind-dispersed seeds. *Nature* **597**, 503–510 (2021).  
<https://doi.org/10.1038/s41586-021-03847-y>

- Received: 06 January 2021
- Accepted: 22 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03847-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

## [Seed-inspired vehicles take flight](#)

- E. Farrell Helbling

News & Views 22 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03847-y>.

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

- Article
- [Published: 22 September 2021](#)

# Three-dimensional magnetic stripes require slow cooling in fast-spread lower ocean crust

- [Sarah M. Maher](#) [ORCID: orcid.org/0000-0001-9454-0099<sup>1</sup>](#),
- [Jeffrey S. Gee<sup>1</sup>](#),
- [Michael J. Cheadle<sup>2</sup>](#) &
- [Barbara E. John<sup>2</sup>](#)

*Nature* volume **597**, pages 511–515 (2021)

- 533 Accesses
- 26 Altmetric
- [Metrics details](#)

## Subjects

- [Geophysics](#)
- [Palaeomagnetism](#)
- [Structural geology](#)
- [Tectonics](#)

## Abstract

Earth's magnetic field is recorded as oceanic crust cools, generating lineated magnetic anomalies that provide the pattern of polarity reversals

for the past 160 million years<sup>1</sup>. In the lower (gabbroic) crust, polarity interval boundaries are proxies for isotherms that constrain cooling and hence crustal accretion. Seismic observations<sup>2,3,4</sup>, geospeedometry<sup>5,6,7</sup> and thermal modelling<sup>8,9,10</sup> of fast-spread crust yield conflicting interpretations of where and how heat is lost near the ridge, a sensitive indicator of processes of melt transport and crystallization within the crust. Here we show that the magnetic structure of magmatically robust fast-spread crust requires that crustal temperatures near the dike–gabbro transition remain at approximately 500 degrees Celsius for 0.1 million years. Near-bottom magnetization solutions over two areas, separated by approximately 8 kilometres, highlight subhorizontal polarity boundaries within 200 metres of the dike–gabbro transition that extend 7–8 kilometres off-axis. Oriented samples with multiple polarity components provide direct confirmation of a corresponding horizontal polarity boundary across an area approximately one kilometre wide, and indicate slow cooling over three polarity intervals. Our results are incompatible with deep hydrothermal cooling within a few kilometres of the axis<sup>2,7</sup> and instead suggest a broad, hot axial zone that extends roughly 8 kilometres off-axis in magmatically robust fast-spread ocean crust.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

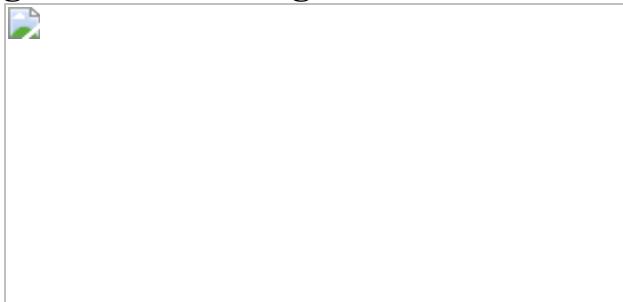
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

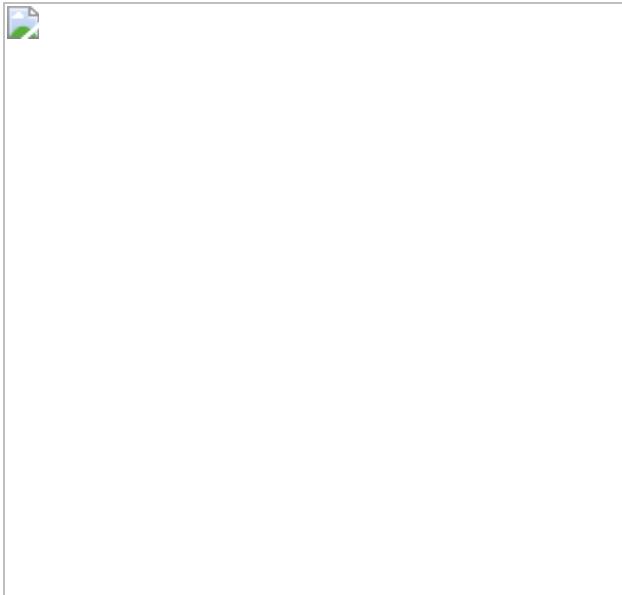
**Fig. 1: Sea surface magnetization solution draped over shaded bathymetry for Pito Deep.**



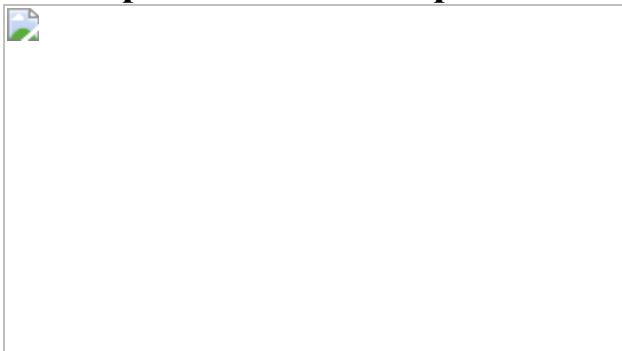
**Fig. 2: Seafloor magnetization and sample locations at Pito Deep.**



**Fig. 3: Magnetic polarity of oriented gabbroic samples ( $n = 229$ ) in area B.**



**Fig. 4: Thermal structure of fast-spread crust forming at the ridge axis before exposure at Pito Deep at 3.330 Ma from our data.**



## Data availability

Near-bottom magnetic data are available at

<https://www.ngdc.noaa.gov/trackline/request/?>

surveyTypes=All%20Parameters&surveyIds=SENTRY418,SENTRY419,SENTRY420,SENTRY421,SENTRY422,SENTRY423,SENTRY424,SENTRY425,SENTRY428 and bathymetry data at

[https://www.ngdc.noaa.gov/auvs/sentry/AT37-08\\_sentry\\_mb.html](https://www.ngdc.noaa.gov/auvs/sentry/AT37-08_sentry_mb.html). Thermal demagnetization data for samples are archived at

<https://doi.org/10.7288/V4/MAGIC/17051>. High-resolution bathymetry

surveys of the East Pacific Rise are from

<https://doi.org/10.26022/IEDA/329855>. Source data are provided with this

paper.

# Code availability

Code available upon request from the corresponding author.

## References

1. 1.

Ogg, J. G. in *The Geologic Time Scale* 85–113 (Elsevier, 2012).

2. 2.

Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9°30'N. *J. Geophys. Res.* **105**, 23537–23555 (2000).

3. 3.

Han, S. et al. Architecture of on- and off-axis magma bodies at EPR 9°37–40'N and implications for oceanic crustal accretion. *Earth Planet. Sci. Lett.* **390**, 31–44 (2014).

4. 4.

Carbotte, S., Canales, J. P., Neimovic, M., Carton, H. & Mutter, J. Recent seismic studies at the East Pacific Rise 8°20'-10°10'N and Endeavour Segment: Insights into mid-ocean ridge hydrothermal and magmatic processes. *Oceanography* **25**, 100–112 (2012).

5. 5.

VanTongeren, J. A., Kelemen, P. B. & Hanghøj, K. Cooling rates in the lower crust of the Oman ophiolite: Ca in olivine, revisited. *Earth Planet. Sci. Lett.* **267**, 69–82 (2008).

6. 6.

Faak, K. & Gillis, K. M. Slow cooling of the lowermost oceanic crust at the fast-spreading East Pacific Rise. *Geology* **44**, 115–118 (2016).

7. 7.

Sun, C. & Lissenberg, C. J. Formation of fast-spreading lower oceanic crust as revealed by a new Mg–REE coupled geospeedometer. *Earth Planet. Sci. Lett.* **487**, 165–178 (2018).

8. 8.

Morton, J. L. & Sleep, N. H. A mid-ocean ridge thermal model: constraints on the volume of axial hydrothermal heat flux. *J. Geophys. Res.* **90**, 11,345–11,353 (1985).

9. 9.

Maclennan, J., Hulme, T. & Singh, S. C. Thermal models of oceanic crustal accretion: Linking geophysical, geological and petrological observations. *Geochem. Geophys. Geosyst.* **5**, Q02F25 (2004).

10. 10.

Hasenclever, J. et al. Hybrid shallow on-axis and deep off-axis hydrothermal circulation at fast-spreading ridges. *Nature* **508**, 508–512 (2014).

11. 11.

Seton, M. et al. A global data set of present-day oceanic crustal age and seafloor spreading parameters. *Geochem. Geophys. Geosyst.* **21**, 1–15 (2020).

12. 12.

Quick, J. E. & Denlinger, R. P. Ductile deformation and the origin of layered gabbro in ophiolites. *J. Geophys. Res.* **98**, 14,015–14,027 (1993).

13. 13.

Morgan, J. P. & Chen, Y. J. The genesis of oceanic crust: magma injection, hydrothermal circulation and crustal flow. *J. Geophys. Res.* **98**, 6283–6297 (1993).

14. 14.

Boudier, F., Nicolas, A. & Ildefonse, B. Magma chambers in the Oman ophiolite: fed from the top and the bottom. *Earth Planet. Sci. Lett.* **144**, 239–250 (1996).

15. 15.

Korenaga, J. & Kelemen, P. B. Melt migration through the oceanic lower crust: a constraint from melt percolation modeling with finite solid diffusion. *Earth Planet. Sci. Lett.* **156**, 1–11 (1998).

16. 16.

Marjanović, M. et al. A multi-sill magma plumbing system beneath the axis of the East Pacific Rise. *Nat. Geosci.* **7**, 825–829 (2014).

17. 17.

Canales, J. P. et al. Network of off-axis melt bodies at the East Pacific Rise. *Nat. Geosci.* **5**, 279–283 (2012).

18. 18.

Coogan, L. A., Jenkin, G. R. T. & Wilson, R. N. Constraining the cooling rate of the lower oceanic crust: a new approach applied to the Oman ophiolite. *Earth Planet. Sci. Lett.* **199**, 127–146 (2002).

19. 19.

Tivey, M. A. Vertical magnetic structure of ocean crust determined from near-bottom magnetic field measurements. *J. Geophys. Res.* **101**, 20275–20296 (1996).

20. 20.
- Macdonald, K. C., Miller, S. P., Luyendyk, B. P., Atwater, T. M. & Shure, L. Investigation of a Vine-Matthews magnetic lineation from a submersible: the source and character of marine magnetic anomalies. *J. Geophys. Res.* **88**, 3403–3418 (1983).
21. 21.
- Cande, S. C. & Kent, D. V. Constraints imposed by the shape of marine magnetic anomalies on the magnetic source. *J. Geophys. Res.* **81**, 4157–4162 (1976).
22. 22.
- Wilson, D. S. & Hey, R. N. The Galapagos axial magnetic anomaly: evidence for the Emperor Event within the Brunhes and for a two-layer magnetic source. *Geophys. Res. Lett.* **8**, 1051–1054 (1981).
23. 23.
- Maher, S. M., Gee, J. S., Doran, A. K., Cheadle, M. J. & John, B. E. Magnetic structure of fast-spread oceanic crust at Pito Deep. *Geochem. Geophys. Geosyst.* **21**, 1–18 (2020).
24. 24.
- Naar, D. F. & Hey, R. N. Tectonic evolution of the Easter Microplate. *J. Geophys. Res.* **96**, 7961–7993 (1991).
25. 25.
- Naar, D. F., Martinez, F., Hey, R. N., Reed, T. B., IV & Stein, S. Pito Rift: how a large-offset rift propagates. *Mar. Geophys. Res.* **13**, 287–309 (1991).
26. 26.

Schouten, H., Klitgord, K. D. & Gallo, D. G. Edge-driven microplate kinematics. *J. Geophys. Res.* **98**, 6689–6701 (1993).

27. 27.

Karson, J. A. et al. Nested-scale investigation of tectonic windows into super-fast spread crust exposed at the Pito Deep Rift, Easter Microplate, SE Pacific. *InterRidge News* **14**, 5–8 (2005).

28. 28.

Horst, A. J., Varga, R. J., Gee, J. S. & Karson, J. A. Paleomagnetic constraints on deformation of superfast-spread oceanic crust exposed at Pito Deep Rift. *J. Geophys. Res.* **116**, B12103 (2011).

29. 29.

Perk, N. W., Coogan, L. A., Karson, J. A., Klein, E. M. & Hanna, H. D. Petrology and geochemistry of primitive lower oceanic crust from Pito Deep: implications for the accretion of the lower crust at the Southern East Pacific Rise. *Contrib. Mineral. Petrol.* **154**, 575–590 (2007).

30. 30.

Cheadle, M. J. et al. Fast-spread lower ocean crust is more complicated than traditionally thought: insights from in-situ crust at Pito Deep. In *AGU Fall Meeting* (AGU, 2019);  
[<https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/542504>](https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/542504)

31. 31.

Gee, J. & Meurer, W. P. Slow cooling of middle and lower oceanic crust inferred from multicomponent magnetizations of gabbroic rocks from the Mid-Atlantic Ridge south of the Kane fracture zone (MARK) area. *J. Geophys. Res.* **107**, EPM 3-1–EPM 3-18 (2002).

32. 32.

Pullaiah, G., Irving, E., Buchan, K. L. & Dunlop, D. J. Magnetization changes caused by burial and uplift. *Earth Planet. Sci. Lett.* **28**, 133–143 (1975).

33. 33.

Benjamin, S. B. & Haymon, R. M. Hydrothermal mineral deposits and fossil biota from a young (0.1 Ma) abyssal hill on the flank of the fast spreading East Pacific Rise: evidence for pulsed hydrothermal flow and tectonic tapping of axial heat and fluids. *Geochem. Geophys. Geosyst.* **7**, (2006).

34. 34.

Violay, M. et al. An experimental study of the brittle-ductile transition of basalt at oceanic crust pressure and temperature conditions. *J. Geophys. Res. Solid Earth* **117**, B03213 (2012).

35. 35.

Brown, T. C. et al. Textural character of gabbroic rocks from Pito Deep: a record of magmatic processes and the genesis of the upper plutonic crust at fast-spreading mid-ocean ridges. *J. Petrol.* **60**, 997–1026 (2019).

36. 36.

Chutas, L. A. M. *Structures in Upper Oceanic Crust: Perspectives from Pito Deep and Iceland* (Duke Univ., 2007).

37. 37.

Tauxe, L., Kylstra, N. & Constable, C. Bootstrap statistics for paleomagnetic data. *J. Geophys. Res.* **96**, 11723–11740 (1991).

38. 38.

Watson, G. S. Large sample theory of the Langevin distribution. *J. Stat. Plan. Infer.* **8**, 245–256 (1983).

39. 39.

Olsen, N. et al. Calibration of the Ørsted vector magnetometer. *Earth Planets Space* **55**, 11–18 (2003).

40. 40.

Kirschvink, J. L. The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. Int.* **62**, 699–718 (1980).

41. 41.

Gee, J., Staudigel, H., Tauxe, L., Pick, T. & Gallet, Y. Magnetization of the La Palma Seamount Series: implications for seamount paleopoles. *J. Geophys. Res.* **98**, 11743–11767 (1993).

42. 42.

Néel, L. Théorie du traînage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Ann. Geophys.* **5**, 99–136 (1949).

43. 43.

Nagy, L. et al. Stability of equidimensional pseudo-single-domain magnetite over billion-year timescales. *Proc. Natl Acad. Sci. USA* **114**, 10356–10360 (2017).

44. 44.

Pariso, J. E. & Johnson, H. P. Do lower crustal rocks record reversals of the Earth's magnetic field? Magnetic petrology of oceanic gabbros from Ocean Drilling Program Hole 735B. *J. Geophys. Res.* **98**, 16013–16032 (1993).

## Acknowledgements

We thank the AUV *Sentry* and ROV *Jason* teams, the crew of the RV *Atlantis*, the science party for data acquisition, and M. Gess, S. Slead and M. Mijum for their help with sample processing. Support for this work was provided through NSF grant numbers OCE-1459387 (J.S.G.) and OCE-1459462 (M.J.C. and B.E.J.).

## Author information

### Affiliations

1. Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA

Sarah M. Maher & Jeffrey S. Gee

2. Department of Geology and Geophysics, University of Wyoming, Laramie, WY, USA

Michael J. Cheadle & Barbara E. John

### Contributions

J.S.G., M.J.C. and B.E.J. designed the experiment and all authors carried out sample and data collection. S.M. and J.S.G. completed the data processing and analysis. S.M.M., J.S.G., M.J.C. and B.E.J. wrote the paper.

### Corresponding author

Correspondence to Sarah M. Maher.

### Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Suzanne Carbotte and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Topographic profiles across the East Pacific Rise.](#)

a) Profiles shown in map view near Pito Deep (white box). b) Flowline profiles plotted as a function of distance from the spreading centre (0 km). Gray arrows indicate the locations of larger offset faults. Lower two profiles are from sea surface swath bathymetry; remaining profiles show high-resolution bathymetry from near-bottom survey

[Source data.](#)

### [Extended Data Fig. 2 Detailed view of area A near-bottom magnetization solution and rock samples.](#)

The dashed black line indicates 4050 mbsl elevation, and the bold black line indicates the interpreted location of dike/gabbro transition as defined by samples<sup>23</sup>

[Source data.](#)

### [Extended Data Fig. 3 Detailed view of area B near bottom magnetic inversion and rock samples.](#)

The bold black line shows the location of the dike/gabbro boundary which is derived from Brown et al.<sup>35</sup>. The dashed line indicates the location of a down-dropped block. The dark and light grey lines outline the 3475 and 3800 mbsl contours, respectively

[Source data.](#)

### **Extended Data Fig. 4 Crossover-corrected sea surface anomaly data.**

The background (grey) shows bathymetry near Pito Deep, and the black box highlights the region shown in Fig. 1. The crossover-corrected anomaly data has a root mean square misfit of 29.7 nT, and was used to generate a sea surface magnetic inversion following the process outlined in Supporting Information in Maher et al.<sup>23</sup>

[Source data.](#)

### **Extended Data Fig. 5 Orientation methods for determining strike and dip.**

The APS 544 miniature orientation sensor is placed flush against the rock face to determine strike and dip. The sample collection basket mounted in front of *ROV Jason II* is perpendicular to its heading and aids in video estimates of strike when flush with the rock face. Photo credit: copyright Woods Hole Oceanographic Institute, courtesy Mike Cheadle of the University of Wyoming.

### **Extended Data Fig. 6 Equal area plot showing a representative Monte Carlo distribution ( $n = 280$ ) illustrating effect of orientation uncertainty on uniform initial remanence direction of $040^\circ/-20^\circ$ .**

Black circles show sample remanence directions with  $20^\circ$  uncertainty in strike and dip. Open (closed) symbols are upper (lower) hemisphere. The

red circle with radius of 24° indicates 1 sigma uncertainty about the mean direction.

[Source data](#)

## **Supplementary information**

[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. 6](#)

### **Rights and permissions**

[Reprints and Permissions](#)

### **About this article**

## Cite this article

Maher, S.M., Gee, J.S., Cheadle, M.J. *et al.* Three-dimensional magnetic stripes require slow cooling in fast-spread lower ocean crust. *Nature* **597**, 511–515 (2021). <https://doi.org/10.1038/s41586-021-03831-6>

- Received: 10 December 2020
- Accepted: 16 July 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03831-6>

## 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/s41586-021-03831-6>

- Article
- [Published: 01 September 2021](#)

# How deregulation, drought and increasing fire impact Amazonian biodiversity

- [Xiao Feng](#) [ORCID: orcid.org/0000-0003-4638-3927](#)<sup>1 na1</sup>,
- [Cory Merow](#)<sup>2 na1</sup>,
- [Zhihua Liu](#) [ORCID: orcid.org/0000-0002-0086-5659](#)<sup>3 na1</sup>,
- [Daniel S. Park](#) [ORCID: orcid.org/0000-0003-2783-530X](#)<sup>4,5 na1</sup>,
- [Patrick R. Roehrdanz](#) [ORCID: orcid.org/0000-0003-4047-5011](#)<sup>6 na1</sup>,
- [Brian Maitner](#) [ORCID: orcid.org/0000-0002-2118-9880](#)<sup>2 na1</sup>,
- [Erica A. Newman](#)<sup>7,8 na1</sup>,
- [Brad L. Boyle](#)<sup>7,9</sup>,
- [Aaron Lien](#)<sup>8,10</sup>,
- [Joseph R. Burger](#)<sup>7,8,11</sup>,
- [Mathias M. Pires](#)<sup>12</sup>,
- [Paulo M. Brando](#) [ORCID: orcid.org/0000-0001-8952-7025](#)<sup>13,14,15</sup>,
- [Mark B. Bush](#) [ORCID: orcid.org/0000-0001-6894-8613](#)<sup>16</sup>,
- [Crystal N. H. McMichael](#) [ORCID: orcid.org/0000-0002-1064-1499](#)<sup>17</sup>,
- [Danilo M. Neves](#)<sup>18</sup>,
- [Efthymios I. Nikolopoulos](#)<sup>19</sup>,
- [Scott R. Saleska](#)<sup>7</sup>,
- [Lee Hannah](#)<sup>6</sup>,
- [David D. Breshears](#) [ORCID: orcid.org/0000-0001-6601-0058](#)<sup>10</sup>,
- [Tom P. Evans](#) [ORCID: orcid.org/0000-0003-4591-1011](#)<sup>20</sup>,
- [José R. Soto](#)<sup>10</sup>,
- [Kacey C. Ernst](#)<sup>21</sup> &
- [Brian J. Enquist](#)<sup>7,22 na1</sup>

[Nature](#) volume **597**, pages 516–521 (2021)

- 4419 Accesses
- 435 Altmetric
- [Metrics details](#)

## Subjects

- [Biodiversity](#)
- [Biogeography](#)

## Abstract

Biodiversity contributes to the ecological and climatic stability of the Amazon Basin<sup>1,2</sup>, but is increasingly threatened by deforestation and fire<sup>3,4</sup>. Here we quantify these impacts over the past two decades using remote-sensing estimates of fire and deforestation and comprehensive range estimates of 11,514 plant species and 3,079 vertebrate species in the Amazon. Deforestation has led to large amounts of habitat loss, and fires further exacerbate this already substantial impact on Amazonian biodiversity. Since 2001, 103,079–189,755 km<sup>2</sup> of Amazon rainforest has been impacted by fires, potentially impacting the ranges of 77.3–85.2% of species that are listed as threatened in this region<sup>5</sup>. The impacts of fire on the ranges of species in Amazonia could be as high as 64%, and greater impacts are typically associated with species that have restricted ranges. We find close associations between forest policy, fire-impacted forest area and their potential impacts on biodiversity. In Brazil, forest policies that were initiated in the mid-2000s corresponded to reduced rates of burning. However, relaxed enforcement of these policies in 2019 has seemingly begun to reverse this trend: approximately 4,253–10,343 km<sup>2</sup> of forest has been impacted by fire, leading to some of the most severe potential impacts on biodiversity since 2009. These results highlight the critical role of policy enforcement in the preservation of biodiversity in the Amazon.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

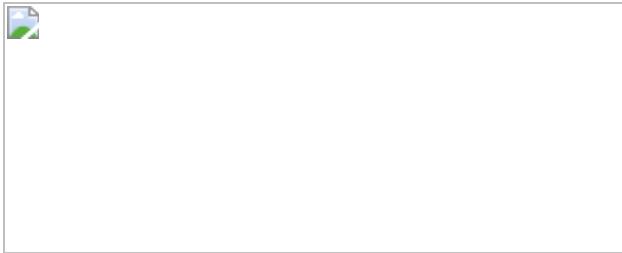
[Rent or Buy](#)

All prices are NET prices.

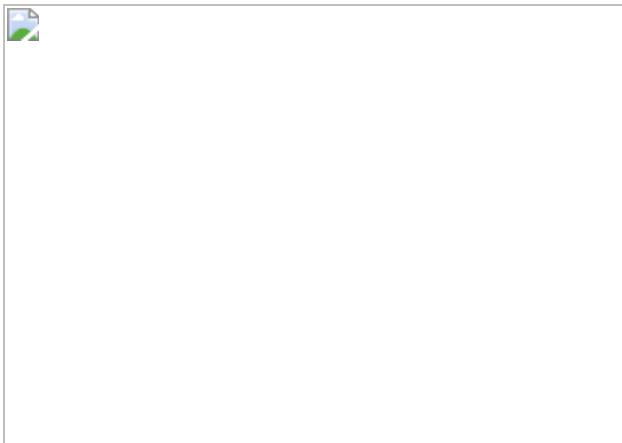
### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

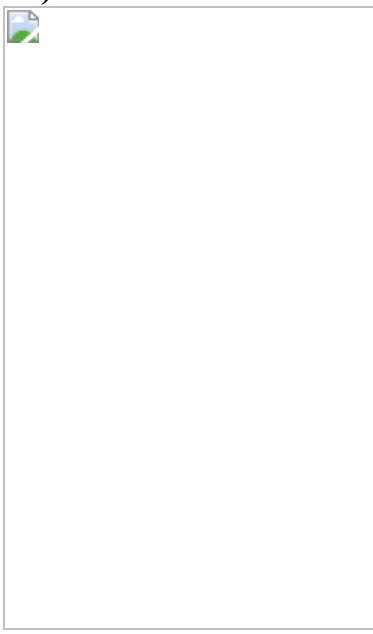
**Fig. 1: Overview of plant and vertebrate species richness and fire-impacted forest in the Amazon Basin.**



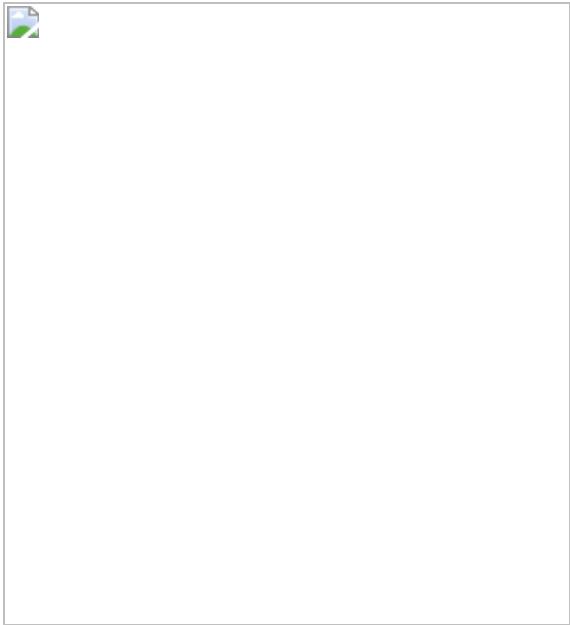
**Fig. 2: Cumulative effects of fire on biodiversity in the Amazon rainforest.**



**Fig. 3: Newly fire-impacted forest in Brazil (based on MODIS burned area).**



**Fig. 4: Newly fire-impacted forest area and the impacts on plant and vertebrate species in Brazil.**



## Data availability

The plant occurrences from the BIEN database are accessible using the RBIEN package (<https://github.com/bmaitner/RBIEN>). The climatic data are accessible from <http://worldclim.org> and the soil data are available from <http://soilgrids.org>. MODIS active fire and burned area products are available at <http://modis-fire.umd.edu>. The MODIS Vegetation Continuous Fields data are publicly available from <https://lpdaac.usgs.gov/products/mod44bv006/>. The annual forest loss layers are available from <http://earthenginepartners.appspot.com/science-2013-global-forest>. The plant range maps are accessible at [https://github.com/shandongfx/paper\\_Amazon\\_biodiversity\\_2021](https://github.com/shandongfx/paper_Amazon_biodiversity_2021). The vertebrate range maps are available from <https://www.iucnredlist.org/resources/spatial-data-download>. The SPEI data are available from SPEI Global Drought Monitor (<https://spei.csic.es/map>).

## Code availability

The code to process the remote-sensing data is available at [https://github.com/shandongfx/paper\\_Amazon\\_biodiversity\\_2021](https://github.com/shandongfx/paper_Amazon_biodiversity_2021).

# References

1. 1.

Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).

2. 2.

Oliver, T. H. et al. Biodiversity and resilience of ecosystem functions. *Trends Ecol. Evol.* **30**, 673–684 (2015).

3. 3.

Barlow, J., Berenguer, E., Carmenta, R. & França, F. Clarifying Amazonia's burning crisis. *Glob. Change Biol.* **9**, 1 (2019).

4. 4.

Brando, P. M. et al. The gathering firestorm in southern Amazonia. *Sci. Adv.* **6**, eaay1632 (2020).

5. 5.

IUCN. *IUCN Red List of Threatened Species version 6.2*.  
<https://www.iucnredlist.org/> (2019).

6. 6.

Flores, M. et al. *WWF's Living Amazon Initiative* (Grambs Corporación Gráfica, 2010).

7. 7.

Hubbell, S. P. et al. How many tree species are there in the Amazon and how many of them will go extinct? *Proc. Natl Acad. Sci. USA* **105 Suppl. 1**, 11498–11504 (2008).

8. 8.

Nepstad, D. C., Stickler, C. M., Filho, B. S.- & Merry, F. Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. *Philos. Trans. R. Soc. Lond. B* **363**, 1737–1746 (2008).

9. 9.

Rankin-de-Mérona, J. M. et al. Preliminary results of a large-scale tree inventory of upland rain forest in the Central Amazon. *Acta Amazon.* **22**, 493–534 (1992).

10. 10.

Sakschewski, B. et al. Resilience of Amazon forests emerges from plant trait diversity. *Nat. Clim. Change* **6**, 1032–1036 (2016).

11. 11.

Poorter, L. et al. Biomass resilience of Neotropical secondary forests. *Nature* **530**, 211–214 (2016).

12. 12.

Beisner, B. E., Haydon, D. T. & Cuddington, K. Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382 (2003).

13. 13.

Lovejoy, T. E. & Nobre, C. Amazon tipping point. *Sci. Adv.* **4**, eaat2340 (2018).

14. 14.

Veldman, J. W. Clarifying the confusion: old-growth savannahs and tropical ecosystem degradation. *Philos. Trans. R. Soc. Lond. B* **371**, (2016).

15. 15.

Arruda, D., Candido, H. G. & Fonseca, R. Amazon fires threaten Brazil's agribusiness. *Science* **365**, 1387 (2019).

16. 16.

Ter Steege, H. et al. Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci. Adv.* **1**, e1500936 (2015).

17. 17.

Gomes, V. H. F., Vieira, I. C. G., Salomão, R. P. & ter Steege, H. Amazonian tree species threatened by deforestation and climate change. *Nat. Clim. Change* **9**, 547–553 (2019).

18. 18.

Brando, P. et al. Amazon wildfires: scenes from a foreseeable disaster. *Flora* **268**, 151609 (2020).

19. 19.

Balch, J. K. et al. The susceptibility of southeastern Amazon forests to fire: insights from a large-scale burn experiment. *Bioscience* **65**, 893–905 (2015).

20. 20.

Barlow, J. et al. The critical importance of considering fire in REDD+ programs. *Biol. Conserv.* **154**, 1–8 (2012).

21. 21.

Cochrane, M. A. & Schulze, M. D. Fire as a recurrent event in tropical forests of the eastern Amazon: effects on forest structure, biomass, and species composition. *Biotropica* **31**, 2–16 (1999).

22. 22.

Brando, P. M. et al. Prolonged tropical forest degradation due to compounding disturbances: Implications for CO<sub>2</sub> and H<sub>2</sub>O fluxes. *Glob. Change Biol.* **25**, 2855–2868 (2019).

23. 23.

Barlow, J. & Peres, C. A. Fire-mediated dieback and compositional cascade in an Amazonian forest. *Philos. Trans. R. Soc. Lond. B* **363**, 1787–1794 (2008).

24. 24.

Cochrane, M. *Tropical Fire Ecology: Climate Change, Land Use and Ecosystem Dynamics* (Springer, 2010).

25. 25.

Uhl, C. & Kauffman, J. B. Deforestation, fire susceptibility, and potential tree responses to fire in the eastern Amazon. *Ecology* **71**, 437–449 (1990).

26. 26.

Cochrane, M. A. Fire science for rainforests. *Nature* **421**, 913–919 (2003).

27. 27.

Cochrane, M. A. & Laurance, W. F. Synergisms among fire, land use, and climate change in the Amazon. *Ambio* **37**, 522–527 (2008).

28. 28.

Nepstad, D. C. et al. Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).

29. 29.

Aragão, L. E. O. C. et al. 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. *Nat. Commun.* **9**, 536 (2018).

30. 30.

Nepstad, D. et al. Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *Science* **344**, 1118–1123 (2014).

31. 31.

Hope, M. The Brazilian development agenda driving Amazon devastation. *Lancet Planet. Health* **3**, e409–e411 (2019).

32. 32.

Brown, J. H. On the relationship between abundance and distribution of species. *Am. Nat.* **124**, 255–279 (1984).

33. 33.

Barnagaud, J.-Y. et al. Ecological traits influence the phylogenetic structure of bird species co-occurrences worldwide. *Ecol. Lett.* **17**, 811–820 (2014).

34. 34.

Šímová, I. et al. Spatial patterns and climate relationships of major plant traits in the New World differ between woody and herbaceous species. *J. Biogeogr.* **45**, 895–916 (2018).

35. 35.

Enquist, B. J. et al. The commonness of rarity: Global and future distribution of rarity across land plants. *Sci. Adv.* **5**, eaaz0414 (2019).

36. 36.

May, P. H., Gebara, M. F., de Barcellos, L. M., Rizek, M. B. & Millikan, B. *The Context of REDD+ in Brazil: Drivers, Agents, and Institutions*, 3rd edition, <https://doi.org/10.17528/cifor/006338> (Center for International Forestry Research, 2016).

37. 37.

Neves, D. M., Dexter, K. G., Baker, T. R., Coelho de Souza, F. & Oliveira-Filho, A. T. Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. *Sci. Rep.* **10**, 1188 (2020).

38. 38.

Cadotte, M. W., Cardinale, B. J. & Oakley, T. H. Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl Acad. Sci. USA* **105**, 17012–17017 (2008).

39. 39.

Hopkins, M. J. G. Modelling the known and unknown plant biodiversity of the Amazon Basin. *J. Biogeogr.* **34**, 1400–1411 (2007).

40. 40.

Wilson, E. O. in *Biodiversity* (eds Wilson E. O. & Peter F. M.) Ch. 1 (National Academies Press, 1988).

41. 41.

Brooks, T. M. et al. Habitat loss and extinction in the hotspots of biodiversity. *Conserv. Biol.* **16**, 909–923 (2002).

42. 42.

Gibbs, H. K. et al. Brazil's soy moratorium. *Science* **347**, 377–378 (2015).

43. 43.

Alix-Garcia, J. & Gibbs, H. K. Forest conservation effects of Brazil's zero deforestation cattle agreements undermined by leakage. *Glob. Environ. Change* **47**, 201–217 (2017).

44. 44.

Escobar, H. There's no doubt that Brazil's fires are linked to deforestation, scientists say. *Science* <https://doi.org/10.1126/science.aaz2689> (2019).

45. 45.

Amazon fires: Brazil sends army to help tackle blazes. *BBC News* <https://www.bbc.co.uk/news/world-latin-america-49452789> (24 August 2019).

46. 46.

Marengo, J. A., Tomasella, J., Soares, W. R., Alves, L. M. & Nobre, C. A. Extreme climatic events in the Amazon basin. *Theor. Appl. Climatol.* **107**, 73–85 (2012).

47. 47.

Malhi, Y. et al. Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proc. Natl Acad. Sci. USA* **106**, 20610–20615 (2009).

48. 48.

Swann, A. L. S. et al. Continental-scale consequences of tree die-offs in North America: identifying where forest loss matters most. *Environ. Res. Lett.* **13**, 055014 (2018).

49. 49.

McCoy, T. Amazon fires dropped unexpectedly in September, after summer spike. *Washington Post* [https://www.washingtonpost.com/world/the\\_americas/amazon-fires-](https://www.washingtonpost.com/world/the_americas/amazon-fires-)

[dropped-unexpectedly-in-september-after-spiking-over-the-summer/2019/10/02/4ddc0026-e516-11e9-b403-f738899982d2\\_story.html](https://www.esajournals.org/doi/10.1890/19-0264.1) (2 October 2019).

50. 50.

Moutinho, P., Guerra, R. & Azevedo-Ramos, C. Achieving zero deforestation in the Brazilian Amazon: what is missing? *Elementa* **4**, 000125 (2016).

51. 51.

Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51**, 933–938 (2001).

52. 52.

Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).

53. 53.

Giglio, L., Schroeder, W. & Justice, C. O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **178**, 31–41 (2016).

54. 54.

Giglio, L. *MODIS Collection 6 Active Fire Product User's Guide Revision A* (Univ. Maryland, 2015).

55. 55.

Barlow, J., Lagan, B. O. & Peres, C. A. Morphological correlates of fire-induced tree mortality in a central Amazonian forest. *J. Trop. Ecol.* **19**, 291–299 (2003).

56. 56.

Brando, P. M. et al. Fire-induced tree mortality in a neotropical forest: the roles of bark traits, tree size, wood density and fire behavior. *Glob. Change Biol.* **18**, 630–641 (2012).

57. 57.

Gibbs, H. K. et al. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl Acad. Sci. USA* **107**, 16732–16737 (2010).

58. 58.

Barlow, J. & Peres, C. in *Emerging Threats to Tropical Forests* (eds. Laurance, W. F. & Peres, C. A.) 225–240 (Univ. Chicago Press, 2006).

59. 59.

Barlow, J. et al. Wildfires in bamboo-dominated Amazonian forest: impacts on above-ground biomass and biodiversity. *PLoS ONE* **7**, e33373 (2012).

60. 60.

Gerwing, J. J. Degradation of forests through logging and fire in the eastern Brazilian Amazon. *For. Ecol. Manage.* **157**, 131–141 (2002).

61. 61.

Brando, P. M. et al. Abrupt increases in Amazonian tree mortality due to drought-fire interactions. *Proc. Natl Acad. Sci. USA* **111**, 6347–6352 (2014).

62. 62.

Barlow, J. & Peres, C. A. Avifaunal responses to single and recurrent wildfires in Amazonian forests. *Ecol. Appl.* **14**, 1358–1373 (2004).

63. 63.

Paolucci, L. N., Schoereeder, J. H., Brando, P. M. & Andersen, A. N. Fire-induced forest transition to derived savannas: cascading effects on ant communities. *Biol. Conserv.* **214**, 295–302 (2017).

64. 64.

Roy, D. P. & Kumar, S. S. Multi-year MODIS active fire type classification over the Brazilian Tropical Moist Forest Biome. *Int. J. Digital Earth* **10**, 54–84 (2017).

65. 65.

Giglio, L., Schroeder, W., Hall, J. V. & Justice, C. O. MODIS Collection 6 Active Fire Product User’s Guide Revision B (Univ. Maryland, 2018).

66. 66.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224 (2011).

67. 67.

Chen, Y. et al. Forecasting fire season severity in South America using sea surface temperature anomalies. *Science* **334**, 787–791 (2011).

68. 68.

Giglio, L. et al. Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences* **7**, 1171–1186 (2010).

69. 69.

Justice, C. O. et al. The MODIS fire products. *Remote Sens. Environ.* **83**, 244–262 (2002).

70. 70.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L. & Justice, C. O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **217**, 72–85 (2018).

71. 71.

Nóbrega, C. C., Brando, P. M., Silvério, D. V., Maracahipes, L. & de Marco, P. Effects of experimental fires on the phylogenetic and functional diversity of woody species in a neotropical forest. *For. Ecol. Manage.* **450**, 117497 (2019).

72. 72.

Alencar, A., Nepstad, D. & Diaz, M. C. V. Forest understory fire in the Brazilian Amazon in ENSO and Non-ENSO years: area burned and committed carbon emissions. *Earth Interact.* **10**, 1–17 (2006).

73. 73.

Siegert, F., Ruecker, G., Hinrichs, A. & Hoffmann, A. A. Increased damage from fires in logged forests during droughts caused by El Niño. *Nature* **414**, 437–440 (2001).

74. 74.

Cochrane, M. A. & Laurance, W. F. Fire as a large-scale edge effect in Amazonian forests. *J. Trop. Ecol.* **18**, 311–325 (2002).

75. 75.

Ray, D., Nepstad, D. & Moutinho, P. Micrometeorological and canopy controls of fire susceptibility in a forested Amazon landscape. *Ecol. Appl.* **15**, 1664–1678 (2005).

76. 76.

Silvérrio, D. V. et al. Fire, fragmentation, and windstorms: a recipe for tropical forest degradation. *J. Ecol.* **107**, 656–667 (2019).

77. 77.

Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Modell.* **135**, 147–186 (2000).

78. 78.

Fegraus, E. Tropical Ecology Assessment and Monitoring Network (TEAM Network). *Biodivers. Ecol.* **4**, 287–287 (2012).

79. 79.

Peet, R. K., Lee, M. T., Jennings, M. D. & Faber-Langendoen, D. VegBank: a permanent, open-access archive for vegetation plot data. *Biodivers. Ecol.* **4**, 233–241 (2012).

80. 80.

DeWalt, S. J., Bourdy, G., Chavez de Michel, L. R. & Quenevo, C. Ethnobotany of the Tacana: quantitative inventories of two permanent plots of Northwestern Bolivia. *Econ. Bot.* **53**, 237–260 (1999).

81. 81.

USDA Forest Service. *Forest Inventory and Analysis National Program*, <http://www.fia.fs.fed.us/> (2013).

82. 82.

Wiser, S. K., Bellingham, P. J. & Burrows, L. E. Managing biodiversity information: development of New Zealand's National Vegetation Survey databank. *N. Z. J. Ecol.* **25**, 1–17 (2001).

83. 83.

Anderson-Teixeira, K. J. et al. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* **21**, 528–549 (2015).

84. 84.

Enquist, B. & Boyle, B. SALVIAS – the SALVIAS vegetation inventory database. *Biodivers. Ecol.* **4**, 288 (2012).

85. 85.

GBIF.org. *GBIF Occurrence Download* <https://doi.org/10.15468/dl.yubndf> (2018).

86. 86.

Dauby, G. et al. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).

87. 87.

Arellano, G. et al. A standard protocol for woody plant inventories and soil characterisation using temporary 0.1-ha plots in tropical forests. *J. Trop. For. Sci.* **28**, 508–516 (2016).

88. 88.

O'Connell, B. M. et al. *The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2 (version 6.1)*, <https://doi.org/10.2737/fs-fiadb-p2-6.1> (USDA Forest Service, 2016).

89. 89.

Oliveira-Filho, A. T. *NeoTropTree, Flora arbórea da Região Neotropical: Um Banco de Dados Envolvendo Biogeografia, Diversidade e Conservação*, <http://www.neotrop-tree.info> (Univ. Federal de Minas Gerais, 2017).

90. 90.

Peet, R. K., Lee, M. T., Jennings, M. D. & Faber-Langendoen, D. *VegBank: The Vegetation Plot Archive of the Ecological Society of America*, <http://vegbank.org> (accessed 2013).

91. 91.

Boyle, B. et al. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinf.* **14**, 16 (2013).

92. 92.

Goldsmith, G. R. et al. Plant-O-Matic: a dynamic and mobile guide to all plants of the Americas. *Methods Ecol. Evol.* **7**, 960–965 (2016).

93. 93.

McFadden, I. R. et al. Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic  $\beta$  diversity. *Ecol. Lett.* **22**, 1126–1135 (2019).

94. 94.

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M. & Thiers, B. M. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. Preprint at <https://doi.org/10.7287/peerj.preprints.2615v2> (2016).

95. 95.

Maitner, B. S. et al. The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379 (2017).

96. 96.

Phillips, S. J. & Dudik, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*

31, 161–175 (2008).

97. 97.

Merow, C. & Silander, J. A. A comparison of Maxlike and Maxent for modelling species distributions. *Methods Ecol. Evol.* **5**, 215–225 (2014).

98. 98.

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B. & Anderson, R. P. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38**, 541–545 (2015).

99. 99.

Grubbs, F. E. Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21**, 27–58 (1950).

100. 100.

Komsta, L. outliers: Tests for outliers. R package v.0.14 (2011).

101. 101.

Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

102. 102.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

103. 103.

Mueller-Dombois, D. & Ellenberg, H. *Aims and Methods of Vegetation Ecology* (Wiley, 1974).

104. 104.

Friedman, J., Hastie, T. & Tibshirani, R. glmnet: Lasso and elastic-net regularized generalized linear models. R package v.4.0-2 (2020).

105. 105.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* **190**, 231–259 (2006).

106. 106.

Drake, J. M. Range bagging: a new method for ecological niche modelling from presence-only data. *J. R. Soc. Interface* **12**, 20150086 (2015).

107. 107.

Cardoso, D. et al. Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl Acad. Sci. USA* **114**, 10695–10700 (2017).

108. 108.

Warton, D. I. & Shepherd, L. C. Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Ann. Appl. Stat.* **4**, 1383–1402 (2010).

109. 109.

Renner, I. W. et al. Point process models for presence-only analysis. *Methods Ecol. Evol.* **6**, 366–379 (2015).

110. 110.

Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience* **67**, 534–545 (2017).

111. 111.

Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2016).

112. 112.

Phillips, S. J. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007). *Ecography* **31**, 272–278 (2008).

113. 113.

Merow, C., Smith, M. J. & Silander, J. A. Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069 (2013).

114. 114.

Qiao, H. et al. An evaluation of transferability of ecological niche models. *Ecography* **42**, 521–534 (2019).

115. 115.

Peterson, A. T., Papeş, M. & Soberón, J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Modell.* **213**, 63–72 (2008).

116. 116.

Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).

117. 117.

Jung, M. et al. Areas of global importance for terrestrial biodiversity, carbon, and water. Preprint at <https://doi.org/10.1101/2020.04.16.201444> (2020).

118. 118.

Carlson, C. J. et al. Climate change will drive novel cross-species viral transmission. Preprint at <https://doi.org/10.1101/2020.01.24.918755> (2020).

119. 119.

BirdLife International. *IUCN Red List for Birds* <http://www.birdlife.org> (2019).

120. 120.

Brooks, T. M. et al. Measuring terrestrial area of habitat (AOH) and its utility for the IUCN Red List. *Trends Ecol. Evol.* **34**, 977–986 (2019).

121. 121.

de Area Leão Pereira, E. J., de Santana Ribeiro, L. C., da Silva Freitas, L. F. & de Barros Pereira, H. B. Brazilian policy and agribusiness damage the Amazon rainforest. *Land Use Policy* **92**, 104491 (2020).

122. 122.

Garcia, R. T. After Brazil's summer of fire, the militarization of the Amazon remains. *Foreign Policy* <https://foreignpolicy.com/2019/11/19/militarization-amazon-legacy-brazil-forest-fire-bolsonaro/> (19 November 2019).

123. 123.

Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **23**, 1696–1718 (2010).

124. 124.

Feldpausch, T. R. et al. Amazon forest response to repeated droughts. *Global Biogeochem. Cycles* **30**, 964–982 (2016).

125. 125.

Marin, P.-G., Julio, C. J., Arturo, R.-T. D. & Jose, V.-N. D. Drought and spatiotemporal variability of forest fires across Mexico. *Chin. Geogr. Sci.* **28**, 25–37 (2018).

126. 126.

Adams, H. D. et al. Temperature sensitivity of drought-induced tree mortality portends increased regional die-off under global-change-type drought. *Proc. Natl Acad. Sci. USA* **106**, 7063–7066 (2009).

## Acknowledgements

We acknowledge the herbaria that contributed data to this work: HA, FCO, MFU, UNEX, VDB, ASDM, BPI, BRI, CLF, L, LPB, AD, TAES, FEN, FHO, A, ANSM, BCMEX, RB, TRH, AAH, ACOR, AJOU, UI, AK, ALCB, AKPM, EA, AAU, ALU, AMES, AMNH, AMO, ANA, GH, ARAN, ARM, AS, CICY, ASU, BAI, AUT, B, BA, BAA, BAB, BACP, BAF, BAL, COCA, BARC, BBS, BC, BCN, BCRU, BEREA, BG, BH, BIO, BISH, SEV, BLA, BM, MJG, BOL, CVRD, BOLV, BONN, BOUM, BR, BREM, BRLU, BSB, BUT, C, CAMU, CAN, CANB, CAS, CAY, CBG, CBM, CEN, CEPEC, CESJ, CHR, ENCB, CHRB, CIIDIR, CIMI, CLEMS, COA, COAH, COFC, CP, COL, COLO, CONC, CORD, CPAP, CPUN, CR, CRAI, FURB, CU, CRP, CS, CSU, CTES, CTESN, CUZ, DAO, HB, DAV, DLF, DNA, DS, DUKE, DUSS, E, HUA, EAC, ECU, EIF, EIU, GI, GLM, GMNHJ, K, GOET, GUA, EKY, EMMA, HUAZ, ERA, ESA, F, FAA, FAU, UVIC, FI, GZU, H, FLAS, FLOR, HCIB, FR, FTG, FUEL, G, GB, GDA, HPL, GENT, GEO, HUAA, HUJ, CGE, HAL, HAM, IAC, HAMAB, HAS, HAST, IB, HASU, HBG, IBUG, HBR, IEB, HGI, HIP, IBGE, ICEL, ICN, ILL, SF, NWOSU, HO, HRCB, HRP, HSS, HU, HUAL, HUEFS, HUEM, HUSA, HUT, IAA, HYO, IAN, ILLS, IPRN,

FCQ, ABH, BAFC, BBB, INPA, IPA, BO, NAS, INB, INEGI, INM, MW, EAN, IZTA, ISKW, ISC, GAT, IBSC, UCSB, ISU, IZAC, JBAG, JE, SD, JUA, JYV, KIEL, ECON, TOYA, MPN, USF, TALL, RELC, CATA, AQP, KMN, KMNH, KOR, KPM, KSTC, LAGU, UESC, GRA, IBK, KTU, KU, PSU, KYO, LA, LOMA, SUU, UNITEC, NAC, IEA, LAE, LAF, GMDRC, LCR, LD, LE, LEB, LI, LIL, LINN, AV, HUCP, MBML, FAUC, CNH, MACF, CATIE, LTB, LISI, LISU, MEXU, LL, LOJA, LP, LPAG, MGC, LPD, LPS, IRVC, MICH, JOTR, LSU, LBG, WOLL, LTR, MNHN, CDBI, LYJB, LISC, MOL, DBG, AWH, NH, HSC, LMS, MELU, NZFRI, M, MA, UU, UBT, CSUSB, MAF, MAK, MB, KUN, MARY, MASS, MBK, MBM, UCSC, UCS, JBGP, OBI, BESA, LSUM, FULD, MCNS, ICESI, MEL, MEN, TUB, MERL, CGMS, FSU, MG, HIB, TRT, BABY, ETH, YAMA, SCFS, SACT, ER, JCT, JROH, SBBG, SAV, PDD, MIN, SJSU, MISS, PAMP, MNHM, SDSU, BOTU, MPU, MSB, MSC, CANU, SFV, RSA, CNS, JEPS, BKF, MSUN, CIB, VIT, MU, MUB, MVFA, SLPM, MVFQ, PGM, MVJB, MVM, MY, PASA, N, HGM, TAM, BOON, MHA, MARS, COI, CMM, NA, NCSC, ND, NU, NE, NHM, NHMC, NHT, UFMA, NLH, UFRJ, UFRN, UFS, ULS, UNL, US, NMNL, USP, NMR, NMSU, XAL, NSW, ZMT, BRIT, MO, NCU, NY, TEX, U, UNCC, NUM, O, OCLA, CHSC, LINC, CHAS, ODU, OKL, OKLA, CDA, OS, OSA, OSC, OSH, OULU, OXF, P, PACA, PAR, UPS, PE, PEL, SGO, PEUFR, PH, PKDC, SI, PMA, POM, PORT, PR, PRC, TRA, PRE, PY, QMEX, QCA, TROM, QCNE, QRS, UH, R, REG, RFA, RIOC, RM, RNG, RYU, S, SALA, SANT, SAPS, SASK, SBT, SEL, SING, SIU, SJRP, SMDB, SNM, SOM, SP, SRFA, SPF, STL, STU, SUVA, SVG, SZU, TAI, TAIF, TAMU, TAN, TEF, TENN, TEPB, TI, TKPM, TNS, TO, TU, TULS, UADY, UAM, UAS, UB, UC, UCR, UEC, UFG, UFMT, UFP, UGDA, UJAT, ULM, UME, UMO, UNA, UNM, UNR, UNSL, UPCB, UPNA, USAS, USJ, USM, USNC, USZ, UT, UTC, UTEP, UV, VAL, VEN, VMSL, VT, W, WAG, WII, WELT, WIS, WMNH, WS, WTU, WU, Z, ZSS, ZT, CUVC, AAS, AFS, BHCB, CHAM, FM, PERTH and SAN. X.F., D.S.P., E.A.N., A.L. and J.R.B. were supported by the University of Arizona Bridging Biodiversity and Conservation Science program. Z.L. was supported by NSFC (41922006) and K. C. Wong Education Foundation. The BIEN working group was supported by the National Center for Ecological Analysis and Synthesis, a centre funded by NSF EF-0553768 at the University of California, Santa Barbara, and the State of California. Additional support

for the BIEN working group was provided by iPlant/Cyverse via NSF DBI-0735191. B.J.E., B.M. and C.M. were supported by NSF ABI-1565118. B.J.E. and C.M. were supported by NSF ABI-1565118 and NSF HDR-1934790. B.J.E., L.H. and P.R.R. were supported by the Global Environment Facility SPARC project grant (GEF-5810). D.D.B. was supported in part by NSF DEB-1824796 and NSF DEB-1550686. S.R.S. was supported by NSF DEB-1754803. X.F. and A.L. were partly supported by NSF DEB-1824796. B.J.E. and D.M.N. were supported by NSF DEB-1556651. M.M.P. is supported by the São Paulo Research Foundation (FAPESP), grant 2019/25478-7. D.M.N. was supported by Instituto Serrapilheira/Brazil (Serra-1912-32082). E.I.N. was supported by NSF HDR-1934712. We thank L. López-Hoffman and L. Baldwin for constructive comments.

## Author information

### Author notes

1. These authors contributed equally: Xiao Feng, Cory Merow, Zhihua Liu, Daniel S. Park, Patrick R. Roehrdanz, Brian Maitner, Erica A. Newman, Brian J. Enquist

## Affiliations

1. Department of Geography, Florida State University, Tallahassee, FL, USA

Xiao Feng

2. Eversource Energy Center and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA

Cory Merow & Brian Maitner

3. CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China

Zhihua Liu

4. Department of Biological Sciences, Purdue University, West Lafayette, IN, USA

Daniel S. Park

5. Purdue Center for Plant Biology, Purdue University, West Lafayette, IN, USA

Daniel S. Park

6. The Moore Center for Science, Conservation International, Arlington, VA, USA

Patrick R. Roehrdanz & Lee Hannah

7. Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Erica A. Newman, Brad L. Boyle, Joseph R. Burger, Scott R. Saleska & Brian J. Enquist

8. Arizona Institutes for Resilience, University of Arizona, Tucson, AZ, USA

Erica A. Newman, Aaron Lien & Joseph R. Burger

9. Hardner & Gullison Associates, Amherst, NH, USA

Brad L. Boyle

10. School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

Aaron Lien, David D. Breshears & José R. Soto

11. Department of Biology, University of Kentucky, Lexington, KY, USA

Joseph R. Burger

12. Departamento de Biologia Animal, Universidade Estadual de Campinas, Campinas, Brazil

Mathias M. Pires

13. Department of Earth System Science, University of California, Irvine, Irvine, CA, USA

Paulo M. Brando

14. Woodwell Climate Research Center, Falmouth, MA, USA

Paulo M. Brando

15. Instituto de Pesquisa Ambiental da Amazônia (IPAM), Brasilia, Brazil

Paulo M. Brando

16. Insitute for Global Ecology, Florida Institute of Technology, Melbourne, FL, USA

Mark B. Bush

17. Department of Ecosystem and Landscape Dynamics, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

Crystal N. H. McMichael

18. Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

Danilo M. Neves

19. Department of Mechanical and Civil Engineering, Florida Institute of Technology, Melbourne, FL, USA

Efthymios I. Nikolopoulos

20. School of Geography, Development and Environment, University of Arizona, Tucson, AZ, USA

Tom P. Evans

21. Department of Epidemiology and Biostatistics, College of Public Health, University of Arizona, Tucson, AZ, USA

Kacey C. Ernst

22. The Santa Fe Institute, Santa Fe, NM, USA

Brian J. Enquist

## **Contributions**

X.F. conceived the idea, which was refined by discussion with D.S.P., C.M., B.M., P.R.R., E.A.N., B.L.B., A.L., J.R.B., D.D.B., J.R.S., K.C.E. and B.J.E.; X.F. and Z.L. processed the remote-sensing data; C.M., X.F., B.M., B.L.B., D.S.P. and B.J.E. conducted the analyses of plant data; P.R.R., C.M., B.M., X.F. and D.S.P. conducted the analyses of vertebrate data; X.F., C.M., S.R.S. and E.A.N. processed the drought data; D.S.P., X.F., C.M., P.R.R. and B.M. designed the illustrations with help from B.J.E., D.D.B., K.C.E. and E.A.N.; E.A.N., X.F., and D.S.P. conducted the statistical analyses with help from B.J.E.; X.F., B.J.E., B.M., A.L., J.R.B., D.S.P., C.M., E.A.N., Z.L. and P.R.R. wrote the original draft; all authors contributed to interpreting the results and the editing of manuscript drafts. B.J.E., C.M., K.C.E. and D.D.B. led to the acquisition of the financial support for the project. X.F., C.M., B.M., D.S.P., P.R.R., Z.L., E.A.N. and B.J.E. contributed equally to data, analyses and writing.

## **Corresponding author**

Correspondence to Xiao Feng.

# Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

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

**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 Fire-impacted forest and forest loss in the Amazon Basin.](#)

**a–h**, Visualization of fire-impacted forest (**a, b**), forest loss without fire (**c, d**), fire-impacted forest with forest loss (**e, f**), and fire-impacted forest without forest loss (**g, h**) in the Amazon Basin based on MODIS burned area (left panels) and active fire (right panels). Data in **a–d** are resampled from the 500m (MODIS burned area) or 1 km (MODIS active fire) to 10 km resolution using mean function and thresholded at 0.01 to illustrate the temporal dynamics. Black represents non-forested areas masked out from this study. The cumulative fire-impacted forest is classified into two categories: fire-impacted forest with forest loss (**e, f**) and fire-impacted forest without forest loss (**g, h**). Data in **e–h** are resampled to 10 km using mean function to illustrate the cumulative percentages of impacts.

### [Extended Data Fig. 2 Scatter plot of species' range impacted by fire.](#)

Scatter plot of species' range size in Amazon forest (x-axis) and percentage of total range impacted by fire (red) and forest loss without fire (black) up to 2019 for plants (left panel) and vertebrates (right panel).

### Extended Data Fig. 3 Density plot of species' cumulative range impacted by fire.

Density plot of species' cumulative range impacted by fire. The different colours represent years 2001-2019. The x-axis is log10 transformed.

### Extended Data Fig. 4 Summary of forest impacts in the Amazon Basin.

Areas of forest impact in the Amazon Basin estimated from MODIS burned area (top) and MODIS active fire (bottom).

### Extended Data Fig. 5 Cumulative impacts on biodiversity in the Amazon Basin.

Cumulative effects of forest loss without fire on biodiversity in the Amazon rainforest. In the left panels, the black and grey shading represent the cumulative forest loss without fire based on MODIS burned area and MODIS active fire, respectively. Coloured areas represent the lower and upper bounds of cumulative numbers of **a**, plant and **c**, vertebrate species' ranges impacted. Right panels depict the relationships between the cumulative forest loss without fire (based on MODIS burned area) and cumulative number of **b**, plant and **d**, vertebrate species. Coloured lines represent predicted values of an ordinary least squares linear regression and grey bands define the two-sided 95% confidence interval (two-sided,  $p$  values = 0.00). The silhouette of the tree is from <http://phylopic.org/>; silhouette of the monkey is courtesy of Mathias M. Pires.

### Extended Data Fig. 6 Fire-impacted forest in Brazil.

Newly fire-impacted forest in Brazil (based on MODIS active fire). **a** shows the area of fire-impacted forest not explained by drought conditions.

Different colours represent years from different policy regimes: pre-regulations in light red (mean value in dark red), regulation in grey (mean value in black dashed line), and 2019 in blue. The y-axis represents the difference between actual area and area predicted by drought conditions calibrated by data from regulation years ([Methods](#)). A positive value on the y-axis represents more area than expected, using the regulation years as a baseline. **b** shows a scatter plot of newly fire-impacted forest in Brazil and drought conditions (SPEI); The lines represent the ordinary least squares linear regression between fire-impacted forest and drought conditions for pre-regulation (red) and regulation (black) respectively.

### [Extended Data Fig. 7 Fire-impacted forest in different countries.](#)

The contribution (0–1) of different countries to the newly fire-impacted forest each year based on MODIS active fire (top) and MODIS burned area (bottom).

### [Extended Data Figure 8 Impacts of fire on forest and biodiversity in Brazil.](#)

**a**, Newly fire-impacted forest, **b**, new range impact on plants and **c**, new range impacts on vertebrate species in Brazil each year (based on MODIS active fire) that are not predicted by drought conditions. The colours represent three policy regimes: pre-regulation in red, regulation in grey and 2019 in blue. The y-axis represents the difference between actual value (area or range impacted by fire) and the values predicted by drought conditions calibrated by data from regulation years ([Methods](#)). A positive value on the y-axis represents more area or range impacted by fire than the expectation using the regulation years as a baseline. The dotted lines represent a smooth curve fitted to the values based on the loess method.

[Extended Data Table 1 Summary of fire-impacted forest](#)  
[Extended Data Table 2 Summary of regression analyses](#)

## **Supplementary information**

## **‘Supplementary Discussion’ and ‘This file contains the Supplementary Discussion’**

### **Reporting Summary**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Feng, X., Merow, C., Liu, Z. *et al.* How deregulation, drought and increasing fire impact Amazonian biodiversity. *Nature* **597**, 516–521 (2021). <https://doi.org/10.1038/s41586-021-03876-7>

- Received: 22 November 2019
- Accepted: 04 August 2021
- Published: 01 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03876-7>

### **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

## **Policy, drought and fires combine to affect biodiversity in the Amazon basin**

- Thomas W. Gillespie

News & Views 01 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03876-7>

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

- Article
- [Published: 22 September 2021](#)

# Paths and timings of the peopling of Polynesia inferred from genomic networks

- [Alexander G. Ioannidis](#) [ORCID: orcid.org/0000-0002-4735-7803<sup>1,2</sup>](#) [na1](#),
- [Javier Blanco-Portillo<sup>2</sup>](#) [na1](#),
- [Karla Sandoval<sup>2</sup>](#),
- [Erika Hagelberg<sup>3</sup>](#),
- [Carmina Barberena-Jonas<sup>2</sup>](#),
- [Adrian V. S. Hill](#) [ORCID: orcid.org/0000-0003-0900-9629<sup>4,5</sup>](#),
- [Juan Esteban Rodríguez-Rodríguez<sup>2</sup>](#),
- [Keolu Fox<sup>6</sup>](#),
- [Kathryn Robson<sup>7</sup>](#),
- [Sonia Haoa-Cardinali<sup>8</sup>](#),
- [Consuelo D. Quinto-Cortés<sup>2</sup>](#),
- [Juan Francisco Miquel-Poblete<sup>9</sup>](#),
- [Kathryn Auckland<sup>4</sup>](#),
- [Tom Parks](#) [ORCID: orcid.org/0000-0002-1163-8654<sup>4</sup>](#),
- [Abdul Salam M. Sofro<sup>10</sup>](#),
- [María C. Ávila-Arcos<sup>11</sup>](#),
- [Alexandra Sockell](#) [ORCID: orcid.org/0000-0002-8725-9415<sup>12</sup>](#),
- [Julian R. Homburger<sup>12</sup>](#),
- [Celeste Eng<sup>13</sup>](#),
- [Scott Huntsman<sup>13</sup>](#),
- [Esteban G. Burchard<sup>13</sup>](#),
- [Christopher R. Gignoux<sup>14</sup>](#),

- [Ricardo A. Verdugo](#)<sup>15,16</sup>,
- [Mauricio Moraga](#)<sup>15,17</sup>,
- [Carlos D. Bustamante](#)<sup>12,18</sup>,
- [Alexander J. Mentzer](#) ORCID: [orcid.org/0000-0002-4502-2209](https://orcid.org/0000-0002-4502-2209)<sup>4,19</sup> &
- [Andrés Moreno-Estrada](#) ORCID: [orcid.org/0000-0001-8329-8292](https://orcid.org/0000-0001-8329-8292)<sup>2</sup>

[Nature](#) volume **597**, pages 522–526 (2021)

- 2800 Accesses
- 437 Altmetric
- [Metrics details](#)

## Subjects

- [Biological anthropology](#)
- [Evolutionary genetics](#)
- [Genomics](#)
- [Population genetics](#)
- [Statistical methods](#)

## Abstract

Polynesia was settled in a series of extraordinary voyages across an ocean spanning one third of the Earth<sup>1</sup>, but the sequences of islands settled remain unknown and their timings disputed. Currently, several centuries separate the dates suggested by different archaeological surveys<sup>2,3,4</sup>. Here, using genome-wide data from merely 430 modern individuals from 21 key Pacific island populations and novel ancestry-specific computational analyses, we unravel the detailed genetic history of this vast, dispersed island network. Our reconstruction of the branching Polynesian migration sequence reveals a serial founder expansion, characterized by directional loss of variants, that originated in Samoa and spread first through the Cook Islands (Rarotonga), then to the Society (Tōtaiete mā) Islands (11th century), the western Austral (Tuha'a Pae) Islands and Tuāmotu Archipelago (12th century), and finally

to the widely separated, but genetically connected, megalithic statue-building cultures of the Marquesas (Te Henua ‘Enana) Islands in the north, Raivavae in the south, and Easter Island (Rapa Nui), the easternmost of the Polynesian islands, settled in approximately ad 1200 via Mangareva.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

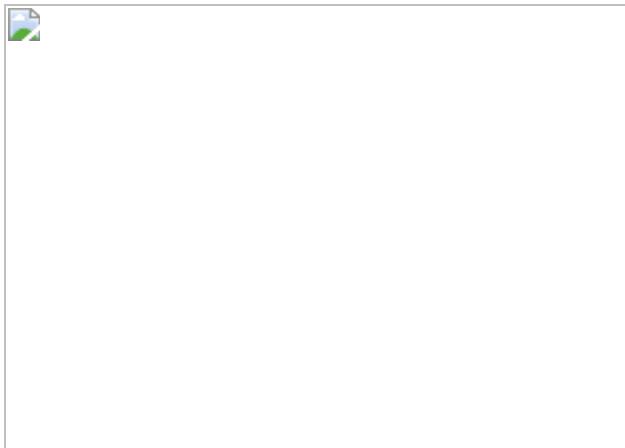
[Rent or Buy](#)

All prices are NET prices.

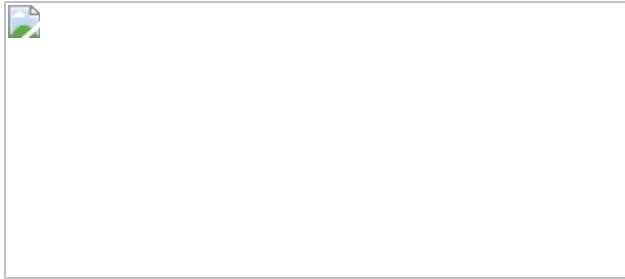
### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: Dimensionality reduction of genetic variation in Pacific Islanders.**



**Fig. 2: Serial bottlenecks and relatedness define the settlement sequence and timings for the Polynesian Islands.**



## Data availability

The samples for this project were collected by the University of Oxford, Stanford University and the University of Chile as part of different studies. SNP data for all newly genotyped individuals are available through a data access agreement to respect the privacy of the participants for the transfer of genetic data from the European Genome-Phenome Archive under accession number EGAS00001005362.

## Code availability

All new techniques described in Methods are available from <https://github.com/AI-sandbox> and all existing software packages and versions used are noted in Methods.

# References

1. 1.

Low, S. *Hawaiki Rising: Hōkūle‘a, Nainoa Thompson, and the Hawaiian Renaissance* (Univ. of Hawaii Press, 2019).

2. 2.

Kirch, P. V. *On the Road of the Winds* (Univ. of California Press, 2017).

3. 3.

Mulrooney, M. A., Bickler, S. H., Allen, M. S. & Ladefoged, T. N. High-precision dating of colonization and settlement in East Polynesia. *Proc. Natl Acad. Sci. USA* **108**, E192–E194 (2011).

4. 4.

Schmid, M. M. E. et al. How 14C dates on wood charcoal increase precision when dating colonization: the examples of Iceland and Polynesia. *Quat. Geochronol.* **48**, 64–71 (2018).

5. 5.

Kahō‘āli‘i Keauokalani, K. Kepelino’s traditions of Hawaii. *Bernice P. Bishop Museum Bulletin* **206** (1932).

6. 6.

Cook, J. *The Journals of Captain James Cook on his Voyages of Discovery* (Cambridge Univ. Press, 1955).

7. 7.

Kirch, P. V. & Green, R. C. *Hawaiki, Ancestral Polynesia* (Cambridge Univ. Press, 2001).

8. 8.
- Minster, R. L. et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat. Genet.* **48**, 1049–1054 (2016).
9. 9.
- Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483 (2009).
10. 10.
- Walworth, M. Eastern Polynesian: the linguistic evidence revisited. *Ocean. Linguist.* **53**, 256–272 (2014).
11. 11.
- Martinsson-Wallin, H., Wallin, P. & Anderson, A. Chronogeographic variation in initial East Polynesian construction of monumental ceremonial sites. *J. Island Coastal Archaeol.* **8**, 405–421 (2013).
12. 12.
- Wilmshurst, J. M., Hunt, T. L., Lipo, C. P. & Anderson, A. J. High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proc. Natl Acad. Sci. USA* **108**, 1815–1820 (2011).
13. 13.
- Spriggs, M. & Anderson, A. Late colonization of east Polynesia. *Antiquity* **67**, 200–217 (1993).
14. 14.
- Hill, A. V. S. et al. Polynesian origins and affinities: globin gene variants in eastern Polynesia. *Am. J. Hum. Genet.* **40**, 453–463 (1987).

15. 15.

Wollstein, A. et al. Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20**, 1983–1992 (2010).

16. 16.

Hudjashov, G. et al. Investigating the origins of eastern Polynesians using genome-wide data from the Leeward Society Isles. *Sci. Rep.* **8**, 1823 (2018).

17. 17.

Skoglund, P. et al. Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).

18. 18.

Posth, C. et al. Language continuity despite population replacement in Remote Oceania. *Nat. Ecol. Evol.* **2**, 731–740 (2018).

19. 19.

McColl, H. et al. The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018).

20. 20.

Emory, K. P. The Tuamotuan creation charts by Paiore. *J. Polynesian Soc.* **48**, 1–29 (1939).

21. 21.

Hunt, T. & Lipo, C. *The Statues that Walked* (Free Press, 2011).

22. 22.

Whyte, A. L. H., Marshall, S. J. & Chambers, G. K. Human evolution in Polynesia. *Hum. Biol.* **77**, 157–177 (2005).

23. 23.

Duncan, R. P., Boyer, A. G. & Blackburn, T. M. Magnitude and variation of prehistoric bird extinctions in the Pacific. *Proc. Natl Acad. Sci. USA* **110**, 6436–6441 (2013).

24. 24.

Steadman, D. W. *Extinction and Biogeography of Tropical Pacific Birds* (Univ. of Chicago Press, 2006).

25. 25.

Kirch, P. V. et al. Human ecodynamics in the Mangareva Islands: a stratified sequence from Nenega-Iti Rock Shelter (site AGA-3, Agakauitai Island). *Archaeol. Oceania* **50**, 23–42 (2015).

26. 26.

Rolett, B. V. Voyaging and interaction in ancient East Polynesia. *Asian Perspect.* **41**, 182–194 (2002).

27. 27.

Handy, E. S. C. *The Native Culture in the Marquesas* (The Bishop Museum, 1923).

28. 28.

Weisler, M. I. et al. Cook Island artifact geochemistry demonstrates spatial and temporal extent of pre-European interarchipelago voyaging in East Polynesia. *Proc. Natl Acad. Sci. USA* **113**, 8150–8155 (2016).

29. 29.

Collerson, K. D. & Weisler, M. I. Stone adze compositions and the extent of ancient Polynesian voyaging and trade. *Science* **317**, 1907–1911 (2007).

30. 30.

Slatkin, M. & Excoffier, L. Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* **191**, 171–181 (2012).

31. 31.

Stephens, M. & Novembre, J. Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* **40**, 646–649 (2008).

32. 32.

Wang, C. et al. Comparing spatial maps of human population-genetic variation using procrustes analysis. *Stat. Appl. Genet. Mol. Biol.* **9**, 13 (2010).

33. 33.

Ioannidis, A. G. et al. Native American gene flow into Polynesia predating Easter Island settlement. *Nature* **583**, 572–577 (2020).

34. 34.

Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).

35. 35.

Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, 2000).

36. 36.

Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

37. 37.

Peter, B. M. & Slatkin, M. Detecting range expansions from genetic data. *Evolution* **67**, 3274–3289 (2013).

38. 38.

Zhan, S. et al. The genetics of monarch butterfly migration and warning colouration. *Nature* **514**, 317–321 (2014).

39. 39.

Lipson, M. et al. Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* **30**, 1788–1802 (2013).

40. 40.

Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).

41. 41.

Leppälä, K., Nielsen, S. V. & Mailund, T. admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017).

42. 42.

Anderson, A. J., Conte, E., Smith, I. & Szabo, K. New excavations at Fa’ahia (Huahine, Society Islands) and chronologies of central East Polynesian colonization. *J. Pac. Arch.* **10**, 1–14 (2019).

43. 43.

Hunt, T. L. & Lipo, C. P. Evidence for a shorter chronology on Rapa Nui (Easter Island). *J. Island Coast. Archaeol.* (2008).

44. 44.

Mulrooney, M. A. An island-wide assessment of the chronology of settlement and land use on Rapa Nui (Easter Island) based on radiocarbon data. *J. Archaeol. Sci.* **40**, 4377–4399 (2013).

45. 45.

Pirazzoli, P. A. & Montaggioni, L. F. Late Holocene sea-level changes in the northwest Tuamotu islands, French Polynesia. *Quat. Res.* **25**, 350–368 (1986).

46. 46.

Di Piazza, A., Di Piazza, P. & Pearthree, E. Sailing virtual canoes across Oceania: revisiting island accessibility. *J. Archaeol. Sci.* **34**, 1219–1225 (2007).

47. 47.

Walworth, M. *The Language of Rapa Iti* (Univ. Hawaii, 2015).

48. 48.

Dickinson, W. Pacific atoll living: how long already and until when. *Geol. Soc. Am. Today* **19**, 4–10 (2009).

49. 49.

Fischer, S. R. Mangarevan doublets: preliminary evidence for proto-southeastern Polynesian. *Ocean. Linguist.* **40**, 112–124 (2001).

50. 50.

Flenley, J. & Bahn, P. *The Enigmas of Easter Island* (Oxford Univ. Press, 2003).

51. 51.

Buck Te Rangi Hīroa, P. H. *Vikings of the Sunrise* (J. B. Lippincott, 1938).

52. 52.

Belbin, G. M. et al. Toward a fine-scale population health monitoring system. *Cell* **184**, 2068–2083.e11 (2021).

53. 53.

Claw, K. G. et al. A framework for enhancing ethical genomic research with Indigenous communities. *Nat. Commun.* **9**, 2957 (2018).

54. 54.

Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

55. 55.

Tyner, C. et al. The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.* **45**, D626–D634 (2017).

56. 56.

Thornton, T. et al. Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91**, 122–138 (2012).

57. 57.

The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

58. 58.

Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

59. 59.

Wickham, H. *ggplot2* (Springer, 2016).

60. 60.

R Core Team. *R: a Language and Environment for Statistical Computing* <https://www.R-project.org/> (2017).

61. 61.

Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinf.* **12**, 246 (2011).

62. 62.

Holmes, S. & Huber, W. *Modern Statistics for Modern Biology* (Cambridge Univ. Press, 2019).

63. 63.

Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).

64. 64.

O’Connell, J. et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* **10**, e1004234 (2014).

65. 65.

D’Arcy, P. The Chinese Pacifics: a brief historical review. *J. Pacific Hist.* **49**, 396–420 (2014).

66. 66.

Browning, S. R. et al. Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* **14**, e1007385 (2018).

67. 67.

Schroeder, H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl Acad. Sci. USA* **115**, 2341–2346 (2018).

68. 68.

Moreno-Estrada, A. et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* **344**, 1280–1285 (2014).

69. 69.

Mazumder, R., Hastie, T. & Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11**, 2287–2322 (2010).

70. 70.

Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).

71. 71.

Skoglund, P. et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).

72. 72.

Moreno-Estrada, A. et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* **9**, e1003925 (2013).

73. 73.

Nguyen, L. H. & Holmes, S. Ten quick tips for effective dimensionality reduction. *PLoS Comp. Biol.* **15**, e1006907 (2019).

74. 74.

Maaten, L. V. D. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

75. 75.

Van Der Maaten, L. Accelerating  $t$ -SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).

76. 76.

McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).

77. 77.

Diaz-Papkovich, A., Anderson-Trocmé, L., Ben-Eghan, C. & Gravel, S. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genet.* **15**, e1008432 (2019).

78. 78.

Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, 2009).

79. 79.

Wittek, P., Gao, S. C., Lim, I. S. & Zhao, L. Somoclu: an efficient parallel library for self-organizing maps. *J. Stat. Softw.* **78**, <https://doi.org/10.18637/jss.v078.i09> (2017).

80. 80.

Nguyen, L. H. & Holmes, S. Bayesian unidimensional scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations. *BMC Bioinf.* **18**, 65–79 (2017).

81. 81.

Peter, B. M. & Slatkin, M. The effective founder effect in a spatially expanding population. *Evolution* **69**, 721–734 (2015).

82. 82.

Pugach, I. et al. The complex admixture history and recent southern origins of Siberian populations. *Mol. Biol. Evol.* **33**, 1777–1795 (2016).

83. 83.

Takahata, N. & Nei, M. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* **110**, 325–344 (1985).

84. 84.

Peter, B. M. Admixture, population structure, and F-statistics. *Genetics* **202**, 1485–1501 (2016).

85. 85.

Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).

86. 86.

Patterson, N. et al. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).

87. 87.

Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).

88. 88.

Nei, M. & Roychoudhury, A. K. Sampling variances of heterozygosity and genetic distance. *Genetics* **76**, 379–390 (1974).

89. 89.

Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Application* (Cambridge Univ. Press, 1997).

90. 90.

Chu, Y. J. & Lui, T. H. On the shortest arborescence of a directed graph. *Science Sinica* **14**, 1396–1400 (1965).

91. 91.

Edmonds, J. Optimum branchings. *J. Res. Natl. Bur. Stand.* **71B**, 233–240 (1967).

92. 92.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).

93. 93.

Huff, C. D. et al. Maximum-likelihood estimation of recent shared ancestry (ERSA). *Genome Res.* **21**, 768–774 (2011).

94. 94.

Baharian, S. et al. The Great Migration and African-American genomic diversity. *PLoS Genet.* **12**, e1006059 (2016).

95. 95.

Gusev, A. et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* **19**, 318–326 (2009).

96. 96.

Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).

97. 97.

Botigué, L. R. et al. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc. Natl Acad. Sci. USA* **110**, 11791–11796 (2013).

98. 98.

Atzmon, G. et al. Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern ancestry. *Am. J. Hum. Genet.* **86**, 850–859 (2010).

99. 99.

Jobling, M., Hurles, M. & Tyler-Smith, C. *Human Evolutionary Genetics* (Garland Science, 2013).

100. 100.

Liang, M. & Nielsen, R. The lengths of admixture tracts. *Genetics* **197**, 953–967 (2014).

101. 101.

Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).

102. 102.

Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, e1001555 (2013).

103. 103.

Deemer, W. L. Jr & Votaw, D. F. Jr Estimation of parameters of truncated or censored exponential distributions. *Ann. Math. Stat.* **26**, 498–504 (1955).

104. 104.

Hill, W. G. & White, I. M. S. Identification of pedigree relationship from genome sharing. *G3 Genes Genom. Genet.* **3**, 1553–1571 (2013).

105. 105.

McVean, G. A. T. et al. The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).

106. 106.

Makarenkov, V. & Lapointe, F.-J. A weighted least-squares approach for inferring phylogenies from incomplete distance matrices. *Bioinformatics* **20**, 2113–2121 (2004).

107. 107.

Fehren-Schmitz, L. et al. Genetic ancestry of Rapanui before and after European contact. *Curr. Biol.* **27**, 3209–3215 (2017).

108. 108.

Crowe, A. *Pathway of the Birds* (Univ. Hawai'i Press, 2018).

109. 109.

Marck, J. C. *Topics in Polynesian Language and Culture History* (The Australian National Univ., 2000).

110. 110.

Niespolo, E. M., Sharp, W. D. & Kirch, P. V. <sup>230</sup>Th dating of coral abraders from stratified deposits at Tangatatau Rockshelter, Mangaia, Cook Islands: implications for building precise chronologies in Polynesia. *J. Archaeol. Sci.* **101**, 21–33 (2019).

111. 111.

Kirch, P. V. *Tangatatau Rockshelter: The Evolution of an Eastern Polynesian Socio-ecosystem* (Cotsen Institute of Archaeology Press, 2017).

112. 112.

Sear, D. A. et al. Human settlement of East Polynesia earlier, incremental, and coincident with prolonged South Pacific drought. *Proc. Natl Acad. Sci. USA* **117**, 8813–8819 (2020).

113. 113.

Kahn, J. G. & Sinoto, Y. Refining the Society Islands cultural sequence: colonisation phase and developmental phase coastal occupation on Mo'orea Island. *J. Polynesian Soc.* **126**, 33 (2017).

114. 114.

Conte, E. & Molle, G. Reinvestigating a key site for Polynesian prehistory: new results from the Hane dune site, Ua Huka (Marquesas). *Archaeol. Oceania* **49**, 121–136 (2014).

115. 115.

Allen, M. S. Marquesan colonisation chronologies and postcolonisation interaction: implications for Hawaiian origins and the ‘Marquesan homeland’ hypothesis. *J. Pac. Arch.* **5**, 1–17 (2014).

116. 116.

Prebble, M. & Wilmshurst, J. M. Detecting the initial impact of humans and introduced species on island environments in Remote Oceania using palaeoecology. *Biol Invasions* **11**, 1529–1556 (2009).

117. 117.

Anderson, A., Kennett, D. J., Culleton, B. J. & Southon, J. in *Taking the High Ground* (eds. Anderson, A. & Kennett, D. J.) 288 (ANU Press, 2012).

118. 118.

Kirch, P. V., Conte, E., Sharp, W. & Nickelsen, C. The Onemea Site (Taravai Island, Mangareva) and the human colonization of Southeastern Polynesia. *Archaeol. Oceania* **45**, 66–79 (2010).

119. 119.

Allen, M. S. & Steadman, D. W. Excavations at the Ureia site, Aitutaki, Cook Islands: preliminary results. *Archaeol. Oceania* **25**, 24–37 (1990).

120. 120.

Matisoo-Smith, E. et al. Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proc. Natl Acad. Sci. USA* **95**, 15145–15150 (1998).

## Acknowledgements

We thank the participants and volunteers who donated DNA samples for this study and the local community leaders and authorities who provided approval and logistics support in the various sampling sites, including T. Teariki (Cook Islands), S. Foliaki (Tonga), N. Tuuau (Samoa), H.-W. Peng (Taiwan), J. Roux (French Polynesia), E. Paoa (Rapa Nui), as well as J. Martinson, D. Weatherall and J. Clegg for their pioneering fieldwork in the region that ultimately led to this work. We thank H. Kane and S. Kahanamoku for critiques and suggestions on the text, as well as M. Spriggs for his comments on the manuscript concerning archaeological dates in Polynesia. We thank the Core Staff at the UCSF Institute for Human Genetics for contributing genotyping capacity, and the Stanford Center for Computational, Evolutionary and Human Genomics (CEHG) for supporting the initial stages of this project. We thank J. Cervantes for IT support and M. Ortega and G. Mireles for technical support at Langebio's Genomics Core Facility at CINVESTAV, Mexico. This work was supported by the George Rosenkranz Prize for Health Care Research in Developing Countries, Mexico's CONACYT Basic Research Program

(grant number CB-2015-01-251380), and the International Center for Genetic Engineering and Biotechnology (ICGEB, Italy) Grant CRP/MEX15-04\_EC (each awarded to A.M.-E.); the Chilean funding programs FONDEF, FONDECYT and CONICYT (grants D10I1007, 1130303 and USA2013-0015, respectively); the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) and a Wellcome Trust Fellowship with reference 106289/Z/14/Z (A.J.M.); together with National Library of Medicine (NLM) training grant T15LM007033 and an American Society of Engineering Education NDSEG Fellowship (each awarded to A.G.I.). Views expressed are those of the author(s) and not necessarily those of the NIHR, the NHS, or the Department of Health.

## Author information

### Author notes

1. These authors contributed equally: Alexander G. Ioannidis, Javier Blanco-Portillo

## Affiliations

1. Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA

Alexander G. Ioannidis

2. National Laboratory of Genomics for Biodiversity (LANGEBIO)—Advanced Genomics Unit (UGA), CINVESTAV, Irapuato, Guanajuato, Mexico

Alexander G. Ioannidis, Javier Blanco-Portillo, Karla Sandoval, Carmina Barberena-Jonas, Juan Esteban Rodríguez-Rodríguez, Consuelo D. Quinto-Cortés & Andrés Moreno-Estrada

3. Department of Biosciences, University of Oslo, Oslo, Norway

Erika Hagelberg

4. Wellcome Centre for Human Genetics, University of Oxford,  
Roosevelt Drive, Oxford, UK

Adrian V. S. Hill, Kathryn Auckland, Tom Parks & Alexander J.  
Mentzer

5. The Jenner Institute, Nuffield Department of Medicine, University of  
Oxford, Oxford, UK

Adrian V. S. Hill

6. Department of Anthropology, University of California San Diego, La  
Jolla, CA, USA

Keolu Fox

7. MRC Weatherall Institute of Molecular Medicine, University of  
Oxford, Oxford, UK

Kathryn Robson

8. Mata Ki Te Rangi Foundation, Hanga Roa, Easter Island, Chile

Sonia Haoa-Cardinali

9. Departamento de Gastroenterología, Facultad de Medicina, Pontificia  
Universidad Católica de Chile, Santiago, Chile

Juan Francisco Miquel-Poblete

10. Department of Biochemistry, Faculty of Medicine, Yayasan Rumah  
Sakit Islam (YARSI) University, Cempaka Putih, Jakarta, Indonesia

Abdul Salam M. Sofro

11. International Laboratory for Human Genome Research (LIIGH),  
UNAM Juriquilla, Queretaro, Mexico

Maria C. Ávila-Arcos

12. Center for Computational, Evolutionary and Human Genomics (CEHG), Stanford University, Stanford, CA, USA

Alexandra Sockell, Julian R. Homburger & Carlos D. Bustamante

13. Program in Pharmaceutical Sciences and Pharmacogenomics, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

Celeste Eng, Scott Huntsman & Esteban G. Burchard

14. Division of Biomedical Informatics and Personalized Medicine, University of Colorado, Denver, CO, USA

Christopher R. Gignoux

15. Human Genetics Program, Institute of Biomedical Sciences, Faculty of Medicine, University of Chile, Santiago, Chile

Ricardo A. Verdugo & Mauricio Moraga

16. Translational Oncology Department, Faculty of Medicine, University of Chile, Santiago, Chile

Ricardo A. Verdugo

17. Department of Anthropology, Faculty of Social Sciences, University of Chile, Santiago, Chile

Mauricio Moraga

18. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

Carlos D. Bustamante

19. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK

Alexander J. Mentzer

## Contributions

A.M.-E. and A.G.I. conceived the study. A.G.I., A.J.M., C.D.B. and A.M.-E. provided overall project supervision and management. A.M.-E., A.G.I., E.H., K.S., A.J.M. and C.R.G. contributed to study design. A.S., C.E., S.H., E.G.B., C.D.B. and A.M.-E. carried out genotyping experiments and quality control. A.G.I., J.B.-P., C. B.-J., J.E.R.-R., C.D.Q.-C., J.R.H. and A.M.-E. analysed the data. A.G.I. developed the analytical methods. A.G.I., J.B.-P. and A.M.-E. interpreted the results. A.M.-E., K.S., E.H., A.V.S.H., J.F.M.-P., K.A., T.P., K.R., M.C.A.-A., A.S., A.S.M.S., C.E., S.H., E.G.B., R.A.V., M.M., A.J.M. and C.D.B. contributed to acquisition of the data. A.G.I. wrote the manuscript, and A.M.-E., E.H., K.F. and S.H.-C. provided feedback on the manuscript.

## Corresponding authors

Correspondence to Alexander G. Ioannidis or Andrés Moreno-Estrada.

## Ethics declarations

## Competing interests

C.D.B. is a member of the scientific advisory boards for Liberty Biosecurity, Personalis, 23andMe Roots into the Future, Ancestry.com, IdentifyGenomics, Genomelink and Etalon, and is a founder of CDB Consulting. C.R.G. owns stock in 23andMe and is member of the scientific advisory board for Encompass Bioscience.

## Additional information

**Peer review information** *Nature* thanks Patrick Kirch, Benjamin Peter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Comparison of genetic and geographic coordinates for European vs. Polynesian samples.](#)

**a**, A principal component analysis of samples from Europe (15 from each nation) is shown to closely fit the geography of Europe. (See Extended Data Fig. 2 for a quantitative comparison.) **b, c**, A principal component analysis (**b**) of samples from Polynesia (with non-Polynesian ancestry masked) is shown not to match the vast geography of the Pacific (**c**), and instead splits out island groups one at a time, reflecting the founder effects that dominate the variance of these populations.

### [Extended Data Fig. 2 Permutation test for fit between genetic and geographic coordinates.](#)

100,000 random permutations of the population labels were created for the European populations' genetic data (blue, left) versus the Polynesian populations (orange, right). For the European populations, out of 100,000 random permutations of the population labels on the genetic PCA, none better fits the geography of Europe (after fitting using a Procrustes analysis<sup>32</sup>), than the correct labels, showing that the genetic coordinates of Europeans fit the geographic coordinates of Europe better than chance every time. However, for the Polynesian data 5% of the random permutations of the labels on the genetic PCA fit the geographic coordinates of the Pacific islands better (after fitting using Procrustes), showing that the genetic data in Polynesia does not fit Polynesia's geography better than random chance. In the box and whiskers plots the mean and upper and lower quartiles of the rms error of the fits of the

random permutations of population labels are indicated by horizontal lines. The fits of the actual population labels are indicated by asterisks.

### **Extended Data Fig. 3 Continuity between ancient and modern Polynesian island populations.**

F3 statistics were computed between ancient Rapanui samples and the Polynesian component from modern samples from each island in our dataset (top)<sup>107</sup>. Indigenous Austronesian language speakers from Taiwan (the Atayal) were used as an outgroup. The ancient Rapanui were found to be the most similar genetically to the modern Rapanui, indicating genetic continuity. A similar comparison was performed between the only other ancient samples from an island in our study, Tonga (bottom)<sup>18</sup>. Again, the modern Tongans appear most similar genetically; however, all islands downstream from Tonga in our inferred settlement path also share the same amount of genetic drift with the ancient Tongan samples (to within one standard error), as they should, since they are all descendants of these ancient Tongan sample according to our settlement reconstruction.

### **Extended Data Fig. 4 Statistics used for settlement path inference.**

All statistics are based on the Polynesian-specific aggregate SNP frequency vectors computed for each island from all sampled individuals. The number ( $n$ ) of individuals used are given for each island in Supplementary Table 1. **a**, Directionality index ( $\psi$ ), used to define sets of potential parent islands, plotted for each island relative to Samoa (equivalent to the top row of the matrix in Fig. 2b). **b**, Average number of pairwise differences ( $\pi$ ), measuring genetic distance and used to select the closest of potential parents, plotted for each island relative to Rapa Nui. **c**, F3 statistic, used to find additional shared genetic drift, plotted for each island relative to Rapa Nui, with Taiwan as an outgroup. Standard errors in **a–c** were determined by a block bootstrap analysis. **d**, Exponential decay constant ( $\lambda$ ) for the Polynesian-specific IBD fragment length distributions between all pairs of individuals from Rapa Nui and each plotted island. The  $\lambda$  values can be used to calculate the number of generations elapsed since each pair of

island populations were joined. Error bars show 95% confidence intervals of the maximum likelihood estimates determined analytically from the Fisher Information.

### **Extended Data Fig. 5 Settlement map with candidate intermediate islands added.**

A reproduction of the map of Fig. 2a showing intermediate islands that are in the settlement path but not in our dataset that are possible candidates for explaining the additional shared drift observed in the corresponding colored settlement branches, that is, genetic drift shared between the child islands but not shared with the parent island. The additional shared drift of the Austral islands (Rimatara and Tubuai) with the Society islands (Tahiti) and Tuamotus (Palliser) beyond what they each share with their parental island (Rarotonga in the Cooks) could indicate that there exists a shared intermediate island in their settlement path that we do not have in our dataset, for instance Mangaia<sup>108</sup>. Geological analyses of ancient tools found on Mangaia (green) have shown that it served as a connection between the Cook islands and remote eastern Polynesia<sup>28</sup>, now uninhabited Nororotu (Maria Atoll) is also believed to have played a role as an intermediary island<sup>108</sup>. Traditional histories give Raiatea (pink) and its surrounding islands a role in the settling of remote eastern Polynesia<sup>108</sup>. Finally, linguistic studies have found connections between Marquesic languages (Marquesas and Mangareva) and the central Tuamotus (orange)<sup>109</sup>. North Marquesas, South Marquesas, and Mangareva share drift with one another beyond what they share with Palliser, the westernmost island group in the Tuamotus, which could indicate that these three populations shared a common settlement path eastward through some of the Tuamotu Archipelago before diverging. Another possible explanation for additional shared drift is the settlement of each child island from a common subpopulation within the parental island, such as from the same clan or village.

### **Extended Data Fig. 6 Effect of phasing errors on IBD dates.**

IBD segments on the island of Rapa Nui were identified between all male X chromosomes. The log of the number of IBD segments (y axis) of a given genetic length (x axis) is plotted (orange; bottom left). The expected exponential decay of IBD segment lengths (linear semilog plot) is seen. The slope of this line ( $-0.161$ ) is the exponential (decay) constant lambda. Since the X chromosome is perfectly phased in men, because it is haploid, the identification of these IBD segments is unaffected by errors introduced through phasing algorithms. To quantify the effect of such errors, synthetic-female individuals were constructed by combining two male X chromosomes to make a diploid pair and to erase the phase information by recording only the genotype. The unphased diploid genotypes so constructed were phased and IBD segments were again identified and plotted (green; bottom right). The difference between the exponential decay constant ( $-0.166$ ) of these statistically phased genotypes and the previous one is seen to be minor (top panel), amounting to three per cent (3.01%), which corresponds to a difference of around 25 years for dates approximately eight hundred years ago (as in Polynesia). Uncertainty in the slope of the lines (equivalent to the uncertainty in the estimate exponential decay constant) is shaded.

### **Extended Data Fig. 7 Polynesian ancestry-specific shared drift ordination plot with principal curve.**

A principal coordinate analysis (PCoA) projection of the pairwise shared drift distances (the Polynesian ancestry-specific outgroup-F3) between each Pacific island population using Taiwan as an outgroup (Supplementary Fig. 12). This PCoA projection uses only the pairwise distance matrix and is fully unsupervised; that is, it does not presuppose that Rapa Nui is a terminal island along some settlement path. Nevertheless, it shows the same ordering as in Supplementary Fig. 9, confirming that Rapa Nui is indeed the terminal island in our dataset along the longest drift path, and confirming the drift ordering along that path. For further confirmation, a principal curve was also fit to the full dimensional space (Supplementary Fig. 12) and then projected into the two-dimensional PCoA space for visualization. The orthogonal projections of each island onto the principal curve are shown as thinner grey lines. This fully unsupervised principal curve confirms the visually apparent path from Island Southeast Asia (Sumatra, far right)

through Samoa, Fiji, Tonga and ending in Raivavae, Mangareva, and Rapa Nui (far left) in that order (cf. migration map in Fig. 2a). This projection of the high dimensional principal curve does not double back on itself, showing that the apparent ordering in this projection is consistent with the original high dimensional ordering. Note that this principal curve is able to fit only one settlement path (the principal one, that is, the longest drift path), which ends in Rapa Nui. Other settlement paths that branch away from this principal (longest) path appear simply as clusters projected onto the principal curve, since islands on those paths share no further drift with the principal path. That is, islands settled along secondary branching paths appear as clusters lying very close to one another along the principal curve. For example, Rapa Iti, which branches off from Rarotonga separately from the main settlement path (Fig. 2a), appears here as coincident with Rarotonga along the principal curve. The eigenvalue for PC1 over the sum of eigenvalues is .997 and for PC 2 is .002 (all eigenvalues are non-negative).

### **Extended Data Table 1 Archaeological and genetic inferred dates for first settlement**

## **Supplementary information**

### **Supplementary Information**

This file contains supplementary text, Figs. 1–33, Tables 1–4, discussion and references.

### **Reporting Summary**

### **Peer Review File**

## **Rights and permissions**

### **Reprints and Permissions**

## About this article

### Cite this article

Ioannidis, A.G., Blanco-Portillo, J., Sandoval, K. *et al.* Paths and timings of the peopling of Polynesia inferred from genomic networks. *Nature* **597**, 522–526 (2021). <https://doi.org/10.1038/s41586-021-03902-8>

- Received: 09 January 2020
- Accepted: 12 August 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03902-8>

### 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

### [Modern Polynesian genomes offer clues to early eastward migrations](#)

- Patrick V. Kirch

News & Views 22 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03902-8>

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

- Article
- Open Access
- [Published: 10 August 2021](#)

# Rare variant contribution to human disease in 281,104 UK Biobank exomes

- [Quanli Wang<sup>1</sup>](#)<sup>nal</sup>,
- [Ryan S. Dhindsa](#) [ORCID: orcid.org/0000-0002-8965-0813](#)<sup>1</sup><sup>nal</sup>,
- [Keren Carss<sup>2</sup>](#)<sup>nal</sup>,
- [Andrew R. Harper](#) [ORCID: orcid.org/0000-0001-5327-0328](#)<sup>2</sup>,
- [Abhishek Nag<sup>2</sup>](#),
- [Ioanna Tachmazidou<sup>2</sup>](#),
- [Dimitrios Vitsios<sup>2</sup>](#),
- [Sri V. V. Deevi](#) [ORCID: orcid.org/0000-0002-0405-4335](#)<sup>2</sup>,
- [Alex Mackay<sup>3</sup>](#),
- [Daniel Muthas<sup>3</sup>](#),
- [Michael Hühn<sup>3</sup>](#),
- [Susan Monkley<sup>3</sup>](#),
- [Henric Olsson](#) [ORCID: orcid.org/0000-0002-5101-8871](#)<sup>3</sup>,
- [AstraZeneca Genomics Initiative](#),
- [Sebastian Wasilewski<sup>2</sup>](#),
- [Katherine R. Smith<sup>2</sup>](#),
- [Ruth March<sup>4</sup>](#),
- [Adam Platt](#) [ORCID: orcid.org/0000-0002-3455-1789](#)<sup>5</sup>,
- [Carolina Haefliger](#) [ORCID: orcid.org/0000-0002-5095-5716](#)<sup>2</sup> &
- [Slavé Petrovski](#) [ORCID: orcid.org/0000-0002-1527-961X](#)<sup>2,6,7</sup>

[Nature](#) volume 597, pages 527–532 (2021)

- 27k Accesses
- 316 Altmetric
- [Metrics details](#)

## Subjects

- [Genetics research](#)
- [Genome-wide association studies](#)
- [Medical genetics](#)
- [Rare variants](#)

## Abstract

Genome-wide association studies have uncovered thousands of common variants associated with human disease, but the contribution of rare variants to common disease remains relatively unexplored. The UK Biobank contains detailed phenotypic data linked to medical records for approximately 500,000 participants, offering an unprecedented opportunity to evaluate the effect of rare variation on a broad collection of traits<sup>1,2</sup>. Here we study the relationships between rare protein-coding variants and 17,361 binary and 1,419 quantitative phenotypes using exome sequencing data from 269,171 UK Biobank participants of European ancestry. Gene-based collapsing analyses revealed 1,703 statistically significant gene–phenotype associations for binary traits, with a median odds ratio of 12.4. Furthermore, 83% of these associations were undetectable via single-variant association tests, emphasizing the power of gene-based collapsing analysis in the setting of high allelic heterogeneity. Gene–phenotype associations were also significantly enriched for loss-of-function-mediated traits and approved drug targets. Finally, we performed ancestry-specific and pan-ancestry collapsing analyses using exome sequencing data from 11,933 UK Biobank participants of African, East Asian or South Asian ancestry. Our results highlight a significant contribution of rare variants to common disease. Summary statistics are publicly available through an interactive portal (<http://azphewas.com/>).

[Download PDF](#)

## Main

The identification of genetic variants that contribute to human disease has facilitated the development of highly efficacious and safe therapeutic agents<sup>3,4,5</sup>. Drug candidates targeting genes with evidence of human disease causality are in fact substantially more likely to be approved<sup>6,7</sup>. Exome sequencing has revolutionized our understanding of rare diseases, uncovering causal rare variants for hundreds of these disorders. However, most efforts for complex human diseases and traits have relied on genome-wide association studies (GWAS), which focus on common variants. Compared with rare variants, common variants tend to confer smaller effect sizes and can be difficult to map to causal genes<sup>8</sup>.

The UK Biobank (UKB) offers an unprecedented opportunity to assess the contribution of both common and rare genetic variation to thousands of human traits and diseases<sup>1,2,9,10,11,12,13</sup>. Testing for the association between rare variants and phenotypes is typically performed at the variant or gene level. Gene-level association tests include collapsing analyses and burden tests, among others<sup>14,15,16,17</sup>. Collapsing analyses are particularly well suited to detect genetic risk for phenotypes driven by an allelic series<sup>16,17,18,19,20,21,22,23</sup> and can provide a clear link between the causal gene and phenotype. Applications of these methods to the first 50,000 UKB exome sequences have indicated an important role of rare variation in complex disease but have also highlighted a need for larger sample sizes<sup>10,11</sup>.

In this study, we performed a genome-wide association study (PheWAS) using exome sequence data from 269,171 UKB participants of European ancestry to evaluate the association between protein-coding variants and 17,361 binary and 1,419 quantitative phenotypes. We first report the diversity of phenotypes and sequence variation present in this cohort. We then performed variant-level and gene-level association tests to identify risk factors across the allele frequency spectrum. Finally, we performed additional collapsing analyses in 11,933 individuals of African, East Asian or South Asian genetic ancestry. Using these results, we implemented a pan-ancestry analysis of 281,104 UKB participants. Altogether, this study comprehensively examines the contribution of rare protein-coding variation to the genetic architecture of complex human diseases and quantitative traits.

## Cohort characteristics

We processed 998 TB of raw exome sequence data from 302,355 UKB participants through a cloud-based bioinformatics pipeline ([Methods](#)). Through stringent quality control, we removed samples with low sequencing quality and from closely related individuals ([Methods](#)). To harmonize variable categorization modes, scaling, and follow-up responses inherent to the phenotype data, we developed PEACOK, a modification of the PHEANT package<sup>24</sup> ([Methods](#)).

We considered 17,361 binary traits and 1,419 quantitative traits, which we categorized into 22 ICD-10-based chapters (Extended Data Fig. [1a, b](#), Supplementary Table [1](#)). We also computed the union of cases across similar phenotypes, resulting in 4,911 union phenotypes ([Methods](#); Supplementary Table [1](#)). The median number of European cases per binary union phenotype was 191 and the median number of individuals tested for quantitative traits was 13,782 (Extended Data Fig. [1c, d](#)). The median number of binary union traits was 25 (Extended Data Fig. [1e](#)).

Approximately 95% of the sequenced UKB participants are of European ancestry (Extended Data Fig. [1f](#)). This affects health-care equity, as the resolution to evaluate

variants across the allele frequency spectrum is proportional to the number of sequenced individuals in a population. For example, individuals from non-European ancestries showed a substantially higher number of rare (minor allele frequency (MAF) < 0.005%), non-synonymous variants in Online Mendelian Inheritance in Man (OMIM) disease-associated genes (Extended Data Fig. 1g). This demonstrates a reduced resolution to accurately estimate lower variant frequencies in non-European populations, as previously observed<sup>25</sup>.

## Identifying protein-truncating variants

Protein-truncating variants (PTVs), which often inactivate proteins, provide direct insight into human biology and disease mechanisms<sup>26,27</sup>. Identifying PTVs that are protective against disease can also offer direct human validation of potential therapeutic targets<sup>5,28</sup>. Among 287,917 participants of any ancestry, we observed that 96% of 18,762 studied genes had at least one heterozygous PTV carrier, 46% had at least one compound heterozygous or homozygous/hemizygous PTV carrier, and 20% had at least one homozygous/hemizygous PTV carrier (Fig. 1a). Only 884 genes (4.7%) had PTVs with a MAF > 0.5% (Fig. 1a), illustrating the power of exome sequencing to detect this important form of variation. Although some have been implicated in human diseases, most common PTVs occur in genes that are less relevant to disease, such as olfactory receptor genes<sup>29</sup>. Focusing on rarer PTVs (MAF < 1%), we observed that 95% of genes had at least one heterozygous PTV carrier, 42% had at least one compound heterozygous or homozygous/hemizygous PTV carrier, and only 15% had at least one homozygous/hemizygous PTV carrier (Extended Data Fig. 2a).

**Fig. 1: Summary of variant-level exome-wide association study results.**

---

 **figure1**

**a**, The number of genes (y axis) with at least the number of PTV carriers (x axis) in 287,917 UKB participants of any ancestry. The dashed line corresponds to the minimum number of carriers typically required to detect individual PTVs with a MAF > 0.5%, that is, 2,873 carriers. Colours represent heterozygous (het.), putative compound heterozygous (comp. het.) and homozygous/hemizygous carriers (recessive). **b**, The MAF distribution of 632 genome-wide significant ExWAS variants associated with binary traits. The inset plot represents the same data limited to variants with MAF < 0.5%. **c**, The distribution of effect sizes for 509 common versus 123 rare (MAF < 0.5%) significant ExWAS variants. The plots in **b** and **c** include variants with the largest effect sizes achieved per gene. **d**, Percentage of ExWAS study-wide significant PTVs ( $n = 24$ ) and missense variants ( $n = 326$ ) that reflect known or novel gene–phenotype relationships. Variants capturing known gene–phenotype relationships were partitioned into those validated in (1) at least one but not all, or (2) all four publicly available databases: FinnGen release r5, OMIM, the GWAS Catalog

(including GWAS Catalog variants within a 50-kb flanking sequence either side of the index variant), and the ClinVar pathogenic/likely pathogenic variant collection.

## Variant-level associations

Exome sequencing enables association tests between phenotypes and protein-coding variants across the allele frequency spectrum. We performed a variant-level exome-wide association study (ExWAS) to test for associations between all 18,780 phenotypes and 2,108,983 variants observed in at least six participants of European ancestry (that is, MAF > 0.001%). We used three genetic models (Methods), equating to 118.8 billion tests. We used a two-sided Fisher's exact test for binary traits and linear regression for quantitative traits. Using a *P* value threshold of  $P \leq 2 \times 10^{-9}$  (Methods) and excluding the MHC region (chromosome 6: 25–35 Mb), we identified 5,193 significant genotype–phenotype associations for binary traits and 41,754 associations for quantitative traits (Supplementary Table 2, 3).

Many of the significant ExWAS signals arose from rare variants (MAF < 0.5%) (Fig. 1b). The rarest significant variant was a frameshift variant in haemoglobin subunit- $\beta$  (*HBB*) associated with thalassaemia (cohort MAF of 0.0013%) (Supplementary Table 3). In the dominant model, rare variants accounted for 26% of statistically significant associations. Furthermore, 21% (227 of 1,088) of binary trait associations and 12% (1,330 of 10,770) of quantitative trait associations identified using the recessive model were not detected using the dominant model. Associations with more common variants have previously been published<sup>9,12</sup>.

The effect sizes of significant rare variant associations were substantially higher than those of common variants (Wilcoxon  $P = 1.1 \times 10^{-57}$ ) (Fig. 1c). While some significant variants are probably in linkage with nearby causal variants, associated PTVs and missense variants often represent the causal variant themselves<sup>26</sup>. Notably, associations for 13% (3 of 24) and 29% (96 of 326) of the significant PTVs and missense variants, respectively, have not been reported in FinnGen release 5, OMIM, ClinVar or the GWAS catalogue<sup>30,31,32</sup> (Fig. 1d, Supplementary Table 4, 5).

We explored how often significant variant-level associations between different variants in the same gene have opposing directions of effect on a phenotype. Among quantitative trait associations with at least five significant non-synonymous variants (MAF < 0.1%) in a particular gene, at least 80% of variants had the same direction of effect (Extended Data Fig. 2b). This is in contrast to disease-associated non-coding variants, which can variably affect the direction of gene expression<sup>33</sup>.

We compared the results of our Fisher's exact tests to regression-based frameworks. While an exact test is robust for rarer variants, regression methods can incorporate

covariates to help to mitigate confounders and are recommended when careful control for confounding cannot be ensured. We performed single-variant association tests across all autosomal variants for 324 Chapter IX binary phenotypes (diseases of the circulatory system; Supplementary Table 29) using SAIGE SPA<sup>12</sup> and REGENIE 2.0.2 (ref. 34), including sex, age, sequencing batch and ten principal components as covariates ([Supplementary Methods](#)). Fisher's exact Phred scores ( $-10 \times \log_{10}(P$  values)) were strongly correlated with those from SAIGE SPA (Pearson's  $r = 0.95$ ) and REGENIE 2.0.2 (Pearson's  $r = 0.94$ ). Fisher's exact  $P$  value statistics were also more conservative for lower frequency variants ( $\text{MAF} \leq 1\%$ ) (Supplementary Table 6). Correlation was higher for signals with a  $P < 1 \times 10^{-8}$  in either Fisher's exact test or SAIGE SPA (Pearson's  $r = 0.99$ ) and Fisher's exact test or REGENIE 2.0.2 (Pearson's  $r = 0.99$ ) (Supplementary Figs. 1, 2, Supplementary Table 6). The median lambda inflation factor  $\lambda_{\text{GC}}$  for the Fisher's exact test was 1.0006 (range: 0.9675–1.0698) compared with a median  $\lambda_{\text{GC}}$  of 0.9953 (range: 0.9372–1.0940) for SAIGE SPA and a median  $\lambda_{\text{GC}}$  of 1.0001 (range: 0.9439–1.0602) for REGENIE 2.0.2 (Supplementary Table 7). Finally, we found that the Fisher's exact test was the most computationally efficient of the three methods (Supplementary Table 6). In this setting, the Fisher's exact test offered a statistically robust and efficient alternative to regression-based approaches, but required careful quality control, case–control harmonization and ancestry pruning before association testing.

## Rare variant collapsing analyses

We also performed gene-level association tests using collapsing analyses. In this approach, the proportion of cases with a qualifying variant was compared with the proportion of controls with a qualifying variant in each gene<sup>16,17,18,19,20,21,22</sup>. We used 12 different sets of qualifying variant filters (models) to test the association between 18,762 genes and 18,780 phenotypes (Methods; Extended Data Table 1), equating to 4.2 billion tests. The models included ten dominant models, one recessive model and one synonymous variant model that served as an empirical negative control (Methods).

Defining a significance threshold posed a challenge due to strong correlation between the 12 models and among the assessed phenotypes. To avoid false claims, we defined two null distributions: an empirical null distribution using the synonymous collapsing model and an  $n$ -of-1 permutation-based null distribution. These approaches independently converged on a study-wide significance threshold of  $P \leq 2 \times 10^{-9}$  (Methods).

We identified 936 significant gene–phenotype relationships for binary traits and 767 for quantitative traits (Fig. 2a, Extended Data Fig. 3, Supplementary Table 8). These associations were enriched for FDA-approved drugs (binary odds ratio (OR): 7.38

(95% CI: 3.71–13.59),  $P = 1.46 \times 10^{-7}$ ; quantitative OR: 3.71 (95% CI: 2.23–5.74),  $P = 7.04 \times 10^{-9}$ ) (Fig. 2b, Extended Data Fig. 4; Methods) and spanned most disease areas and disease-relevant biomarkers (Fig. 2c,d). Many signals were of large effect, with a median OR of 12.4 for binary traits and a median absolute beta of 0.35 for quantitative traits. We also detected several significant genes with putatively protective PTVs, including *APOB* and *PCSK9* (Supplementary Table 9). The median genomic inflation factor ( $\lambda$ ) was 1.002 for binary traits (range: 0.71–1.35) and 1.010 for quantitative traits (range: 0.88–1.37) (Extended Data Fig. 5a). Only 0.76% of the associations from the 191,037 non-recessive collapsing analyses were outside the 0.9–1.1  $\lambda$  range. Our tests were thus highly robust to systematic bias and other sources of inflation. Collectively, these findings provide biological insight into common diseases and substrates for future therapeutic development opportunities.

**Fig. 2: Summary of gene-level collapsing analysis results.**

 figure2

**a**, Gene–phenotype associations for binary traits. For gene–phenotype associations that appear in multiple collapsing models, we display only the association with the strongest effect size. The dashed line represents the genome-wide significant  $P$  value threshold ( $2 \times 10^{-9}$ ). The  $y$  axis is capped at  $-\log_{10}(P) = 50$  and only associations with  $P < 10^{-5}$  were plotted ( $n = 94,208$ ). **b**, Enrichment of FDA-approved drug targets<sup>6,46</sup> among significant binary traits, quantitative traits, OMIM genes and GWAS signals.  $P$  values were generated via two-sided Fisher’s exact test (\* $P < 10^{-5}$ , \*\* $P < 10^{-20}$ ,

\*\*\* $P < 10^{-70}$ ). Exact statistics: binary odds ratio (OR) = 7.38, 95% CI: 3.71–13.59,  $P = 1.5 \times 10^{-7}$ ; quantitative OR = 3.71, 95% CI: 2.28–5.76,  $P = 4.5 \times 10^{-7}$ ; OMIM OR = 5.95, 95% CI: 4.90–7.23,  $P = 1.1 \times 10^{-75}$ ; GWAS OR = 2.68, 95% CI: 2.12–3.32,  $P = 3.6 \times 10^{-23}$ ). Error bars represent 95% CIs. Contingency tables were created using each of the binary ( $n = 195$ ), quantitative ( $n = 395$ ), OMIM ( $n = 3,875$ ) and GWAS ( $n = 10,692$ ) categories, alongside approved targets from Informa Pharmaprojects ( $n = 463$ ).  $P$  values were generated via a two-tailed Fisher's exact test. **c**, Effect sizes for select gene associations per disease area. Genes with the highest OR for a chapter or with OR > 100 are labelled. **d**, Illustration of large effect gene–phenotype associations for select disease-related quantitative traits. FEV1/FVC, forced expiratory volume in 1 s/forced vital capacity ratio; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Dashed line corresponds to a beta of 0.

Collapsing models focused on PTVs explained 80% of binary and 55% of quantitative associations. Remaining signals emerged from models that included missense variants. While these results confirm the importance of PTVs, they also emphasize the role of other forms of variation in human disease. We found that using the missense tolerance ratio (MTR) to retain missense variants only in constrained genic sub-regions improved the signal-to-noise ratio. Specifically, 15% (133 of 878) of significant relationships detected via the three MTR-informed models were not detected in analogous models that did not incorporate MTR<sup>35</sup>. Moreover, for phenotype associations where both MTR and non-MTR versions of a model achieved significance, effect sizes were significantly higher in the MTR-informed versions (Mann–Whitney test  $P = 0.006$ ; Supplementary Fig. 3). Thus, MTR appears to effectively prioritize putatively functional missense variation in collapsing analyses of complex disease.

Most binary phenotype associations were supported by OMIM or were annotated as pathogenic/likely pathogenic in ClinVar (88.6%), indicating that we robustly captured high-confidence signals (Supplementary Table 10). However, we also identified rare variant associations with phenotypes beyond those reported in OMIM (Supplementary Table 10). For example, 12.1% of the European cohort carried at least one of the 373 distinct filaggrin (*FLG*) PTVs identified. These individuals had a significantly higher risk of well-known associations, including dermatitis ( $P = 5.1 \times 10^{-95}$ ; OR: 1.96 (95% CI: 1.84–2.08)) and asthma ( $P = 3.1 \times 10^{-32}$ ; OR: 1.24 (95% CI: 1.19–1.28))<sup>36</sup>, but were also at risk of under-recognized associations, such as melanoma ( $P = 4.7 \times 10^{-13}$ ; OR: 1.21 (95% CI: 1.15–1.27))<sup>37</sup> and basal cell carcinoma ( $P = 9.9 \times 10^{-10}$ ; OR: 1.19 (95% CI: 1.12–1.25))<sup>38</sup>. Concomitant increases in the levels of vitamin D ( $P = 2.3 \times 10^{-131}$ ;  $\beta$ : 0.15 (95% CI: 0.14–0.16))<sup>39</sup> suggest that the increased risk of skin cancer may be attributable to increased sensitivity to ultraviolet B radiation. This interrogation offers one example of how this genome-wide resource can uncover a

wide spectrum of phenotypes associated with rare variation in any protein-coding gene.

Although our pipeline was tuned for detecting germline variants, we identified seven genes that were significantly associated with haematological malignancies, driven by qualifying variants that appeared to be somatic (Supplementary Tables 11, 12, Supplementary Fig. 4). This supports the potential of blood-based sequencing to yield insight into blood cancer genomes via incidentally detected somatic variants<sup>40</sup>.

Compared with two smaller UKB PheWAS studies<sup>10,11</sup>, we observed a 1.2-fold and 5.6-fold increase, respectively, in statistically significant gene–trait associations using the same first tranche of 50K UKB data, attributable to both the depth of outcomes studied and differences in methodologies (Extended Data Fig. 5b). Increasing the cohort size from 50,000 to the current full dataset led to an 18-fold increase in statistically significant gene–trait associations using our collapsing method (Extended Data Fig. 5c). Incorporating updated phenotypic data from the July 2020 release resulted in a 24-fold increase in significant associations compared with the 50K data (Extended Data Fig. 5c).

Among significant collapsing analysis signals, only 17% (125 of 724) of binary associations and 58% (446 of 767) of quantitative associations were detectable via ExWAS (Supplementary Table 13A). Conversely, most rare PTV ExWAS associations were detected via collapsing analyses, although the rates were lower for rare missense variants (Supplementary Table 13B). Thus, collapsing analyses can identify rare variant associations that are currently undetectable via single-variant-based approaches (Supplementary Table 14).

## Pan-ancestry collapsing analysis

The inclusion of individuals from non-European ancestries in genetic analyses is crucial for health-care equity and genetic discovery<sup>41</sup>. Therefore, we performed additional collapsing analyses for each major non-European ancestral group (that is, South Asian ( $n = 5,714$ ), African ( $n = 4,744$ ) and East Asian ( $n = 1,475$ )). We limited each PheWAS to binary traits with at least five cases in the population and quantitative traits with at least five qualifying variants carriers (Supplementary Table 1).

The only study-wide significant ( $P \leq 2 \times 10^{-9}$ ) binary trait association among the non-European populations was between PTVs in *HBB* and thalassaemia in individuals of South Asian ancestry ( $P = 2.7 \times 10^{-46}$ ; OR = 176.4 (95% CI: 84.1–369.7)) (Supplementary Table 15). We next applied the Cochran–Mantel–Haenszel test to combine the results of the binary trait collapsing analysis across all four studied ancestral groups, including the European population (Methods). This pan-ancestry

PheWAS identified 26 unique study-wide significant gene–phenotype associations that were not significant in the European analyses (Fig. 3a, Extended Data Fig. 6a, Supplementary Table 16). Conversely, 20 gene–phenotype associations that were significant in the European analyses did not reach the study-wide significance threshold in the pan-ancestry analysis.

**Fig. 3: Pan-ancestry collapsing analysis.**



**a, b,** The change in Phred scores between the pan-ancestry and European-only analyses for 46,769 binary associations (a) and 39,541 quantitative associations (b) stratified by chapter. For gene–phenotype associations that appear in multiple collapsing models, we display only those with the lowest  $P$  value. The green dots indicate associations that were not significant in the European analysis but were significant in the combined analysis. The orange dots represent associations that were originally significant in the European-only analysis but became not significant in the combined analysis. In both figures, the  $y$  axis is capped at  $\Delta\text{Phred} = 40$  (equivalent to a  $P$  value change of 0.0001).

We analysed 1,419 quantitative traits in a linear regression model including individuals of all major ancestral groups, including Europeans (Supplementary Table 1). This model included categorical ancestral groups, the top five ancestry principal components, age and sex as covariates (Methods). We identified 59 significant gene–quantitative trait associations that were originally not significant in the European analyses (Fig. 3b, Extended Data Fig. 6b). These included associations between rare variants in *OCA2* and a younger age of wearing glasses ( $P = 4.7 \times 10^{-10}$ ;  $\beta: -0.45$  (95% CI:  $-0.60$  to  $-0.31$ )), *ASGR1* and reduced low-density lipoprotein cholesterol ( $P = 1.7 \times 10^{-9}$ ;  $\beta: -0.26$  (95% CI:  $-0.34$  to  $-0.17$ )), and others (Supplementary Table 17). In addition, 46 unique associations between genes and quantitative traits, originally significant in the European analyses, were not significant in the combined analysis.

## Discussion

We performed a PheWAS using exome sequences of 269,171 UKB participants of European ancestry combined with records of 18,780 phenotypes, followed by a pan-ancestry analysis that incorporated an additional 11,933 UKB participants of African, East Asian and South Asian ancestries. In total, we identified 46,837 variant-level and 1,703 gene-level statistically significant relationships. Many associations were previously known, but others were either new or associated with phenotype expansions. We also found that these associations were significantly enriched for targets of US Food and Drug Administration (FDA)-approved drugs, reinforcing the importance of human genetics in target identification. When followed up with functional investigation to understand biological mechanisms, these results can help to improve the efficiency of pharmaceutical pipelines, contribute towards safety assessments and reveal repositioning opportunities<sup>7,42</sup>.

Our variant-level association tests detected rare variant associations that are not frequent enough to be captured by microarray-based studies (that is, as rare as MAF = 0.0012%). Our gene-level collapsing analyses evaluated the aggregate effect of private-to-rare functional variants, 83% of which were not detected in single-variant tests for binary traits. Among gene-level signals for which an individual variant also achieved significance, we found examples where both common and rare risk variants in these genes contributed to disease burden. This is consistent with previous work demonstrating that common and rare PTVs in *FLG* have similar effect sizes for the risk of early asthma<sup>43</sup>.

We used a Fisher's exact test framework for our variant-level and gene-level analyses based on previous success with this approach<sup>16,17,18,19,20,21,22,23</sup>. Limitations of the Fisher's test compared with regression-based approaches<sup>12,34,44,45</sup> include an inability to adjust for covariates. On a subset of traits selected for comparisons, we observed that the Phred scores for significant variants from the Fisher's exact test, SAIGE SPA and REGENIE 2.0.2 were nearly perfectly correlated (Pearson's  $r = 0.99$ ). The Fisher's exact test generated more conservative statistics for rare variants and was associated with increased computational efficiency. Use of the Fisher's exact test requires extremely careful quality control, case-control harmonization and ancestry pruning. In the absence of these measures, it is crucial to correct for such confounders via a regression-based approach. Future work should focus on in-depth benchmarking for these different methods. Regardless of the approach used, it is essential to define an appropriate study-wide significance threshold, which we addressed using *n*-of-1 permutation and an empirical null distribution using a synonymous negative control model.

The predominant representation of European ancestry in human genomics has negative ethical and clinical consequences<sup>25,41</sup>. Smaller sample sizes limited our ability to detect many associations among individual non-European populations. Performing a

combined pan-ancestry PheWAS bolstered the association signal for several binary and quantitative traits. Altogether, these results emphasize the need to establish more diverse biobanks.

The UKB has set an excellent standard for linking genomic and phenotypic data and its dynamic nature will facilitate new opportunities for genetic discovery. In future studies, phenotypes may be refined through combining binary, phenotypic and temporal data. The results of this PheWAS are publicly available (<http://azphewas.com/>), which we anticipate will help to elucidate disease mechanisms, identify phenotypic expansions and enable the development of human genetically validated drugs.

## Methods

### UKB resource

The UKB is a prospective study of approximately 500,000 participants 40–69 years of age at recruitment. Participants were recruited in the UK between 2006 and 2010 and are continuously followed<sup>47</sup>. The average age at recruitment for sequenced individuals was 56.5 years and 54% of the sequenced cohort comprises those of female sex. Participant data include health records that are periodically updated by the UKB, self-reported survey information, linkage to death and cancer registries, collection of urine and blood biomarkers, imaging data, accelerometer data and various other phenotypic end points<sup>1</sup>. All study participants provided informed consent.

### Phenotypes

We studied two main phenotypic categories: binary and quantitative traits taken from the February 2020 data release that was accessed on 27 March 2020 as part of UKB application 26041. To parse the UKB phenotypic data, we developed a modified version of the PHESENT package, which can be located at <https://github.com/astrazeneca-cgr-publications/PEACOK>. The adopted parameters are available in Supplementary Methods and have been previously introduced in PHESENT (<https://github.com/MRCIEU/PHESENT>)<sup>24</sup>.

The PEACOK R package implementation focuses on separating phenotype matrix generation from statistical association tests. It also allows statistical tests to be performed separately on different computing environments, such as on a high-performance computing cluster or an AWS Batch environment. This package introduces additional functionalities, including the ability to generate phenotypes for every node from a tree-like UKB data code (for example, an ICD-10 code) and to run

logistic regression on a binary phenotype with covariates. Various downstream analysis and summarization were performed using R v3.4.3 <https://cran.r-project.org>. R libraries data.table (v1.12.8; <https://CRAN.R-project.org/package=data.table>), MASS (7.3-51.6; <https://www.stats.ox.ac.uk/pub/MASS4/>), tidyverse (1.1.0; <https://CRAN.R-project.org/package=tidyverse>) and dplyr (1.0.0; <https://CRAN.R-project.org/package=dplyr>) were also used.

In total, 44 UKB paths were represented for the binary traits and 49 for the quantitative traits. For UKB tree fields, such as the ICD-10 hospital admissions (field 41202), we studied each leaf individually and studied each subsequent higher-level groupings up to the ICD-10 root chapter as separate phenotypic entities. Furthermore, for the tree-related fields (fields: 20001, 20002, 40001, 40002, 40006 and 41202), we restricted controls to participants who did not have a positive diagnosis for any phenotype contained within the corresponding chapter to reduce potential contamination due to genetically related diagnoses. A minimum of 30 cases were required for a binary trait to be studied.

In addition to studying UKB algorithmically defined outcomes, we constructed a union phenotype for each ICD-10 phenotype. These union phenotypes are denoted by a ‘Union’ prefix and the applied mappings are available in Supplementary Table 1.

In total, we studied 17,361 binary and 1,419 quantitative phenotypes. For all binary phenotypes, we matched controls by sex when the percentage of female cases was significantly different (Fisher’s exact two-sided  $P < 0.05$ ) from the percentage of available female controls. This included sex-specific traits in which, by design, all controls would be same sex as cases. As a result, 10,531 (60.7%) of the binary phenotypes required down sampling of controls to match the case female percentage (Supplementary Table 1). Finally, to allow for more compartmentalized ICD-10 chapter-based analyses, all 18,780 binary and quantitative trait phenotypes were mapped to a single ICD-10 chapter including manual mapping for the non-ICD-10 phenotypes. Chapter mappings are provided in Supplementary Table 1. It is acknowledged that chapter mapping may have the greatest utility for diagnostic, rather than procedural, ICD-10 codes. For procedural codes, genetic associations could be incorrectly interpreted if chapter mappings are relied on. For example, surgical procedures commonly performed for patients with cancer are categorized within the dermatology chapter. Genetic associations reported for these procedures would be categorized within the dermatology chapter, but the underlying disease process is instead most probably reflective of an oncological aetiology.

We subsequently re-analysed the 300Kv1 cohort using the updated Hospital Episode Statistic (HES) and death registry data as released ad hoc by the UKB on July 2020. Among Data-Field 41270 of primary and secondary inpatient diagnoses that contribute to the Union phenotypes, we found on average a 38.1% increase in the number of

cases when comparing the April 2017 refresh to the July 2020 refresh. Throughout this article, we adopt the July 2020 refresh data as the default analysis dataset and refer to this update as the ‘300Kv2’ dataset. The effect on case numbers before and after updating to this release are documented in Supplementary Table 1.

## Sequencing

Whole-exome sequencing data for UKB participants were generated at the Regeneron Genetics Center (RGC) as part of a pre-competitive data generation collaboration between AbbVie, Alnylam Pharmaceuticals, AstraZeneca, Biogen, Bristol-Myers Squibb, Pfizer, Regeneron and Takeda with the UKB<sup>2</sup>. Genomic DNA underwent paired-end 75-bp whole-exome sequencing at Regeneron Pharmaceuticals using the IDT xGen v1 capture kit on the NovaSeq6000 platform. Conversion of sequencing data in BCL format to FASTQ format and the assignments of paired-end sequence reads to samples were based on 10-base barcodes, using bcl2fastq v2.19.0. Exome sequences from 302,355 UKB participants were made available to the Exome Sequencing consortium in December 2019. Initial quality control was performed by Regeneron and included sex discordance, contamination, unresolved duplicate sequences and discordance with microarray genotyping data checks<sup>11</sup>.

## AstraZeneca Centre for Genomics Research (CGR) bioinformatics pipeline

The 302,355 UKB exome sequences were processed at AstraZeneca from their unaligned FASTQ state. A custom-built Amazon Web Services (AWS) cloud compute platform running Illumina DRAGEN Bio-IT Platform Germline Pipeline v3.0.7 was used to align the reads to the GRCh38 genome reference and perform single-nucleotide variant (SNV) and insertion and deletion (indel) calling. SNVs and indels were annotated using SnpEFF v4.3<sup>48</sup> against Ensembl Build 38.92<sup>49</sup>. We further annotated all variants with their genome Aggregation Database (gnomAD) MAFs (gnomAD v2.1.1 mapped to GRCh38)<sup>27</sup>. We also annotated missense variants with MTR and REVEL scores<sup>35,50</sup>.

## Additional quality control

To complement the quality control performed by Regeneron Pharmaceuticals, we passed the 302,355 sequences through our internal bioinformatics pipeline. In addition to what had already been flagged for quality control, we excluded from our analyses 106 (0.035%) sequences that achieved a VerifyBAMID freemix (contamination) level of more than 4%<sup>51</sup>, and an additional five sequences (0.002%) where less than 94.5%

of the consensus coding sequence (CCDS release 22) achieved a minimum of tenfold read depth<sup>52</sup>.

To mitigate a possible increase of variance estimates due to relatedness, we sought to remove related individuals from our analyses. Using exome sequence-derived genotypes for 43,889 biallelic autosomal SNVs located in coding regions as input to the kinship algorithm included in KING v2.2.3<sup>53</sup>, we generated pairwise kinship coefficients for all remaining samples.

We used the ukb\_gen\_samples\_to\_remove() function from the R package ukbtools v0.11.3<sup>54</sup> to choose a subset of individuals within which no pair had a kinship coefficient exceeding 0.0884, equivalent of up to third-degree relatives. For each related pair, this function removes whichever member has the highest number of relatives above the provided threshold, resulting in a maximal set. Through this process, an additional 14,326 (4.74%) sequences were removed from downstream analyses.

After the above quality control steps, there remained 287,917 (95.2%) predominantly unrelated sequences of any genetic ancestry that were available for analyses presented in this work.

## Genetic ancestry

For most of the case–control cohort analyses, we restricted the statistical tests to include a homogeneous European genetic ancestry test cohort. We predicted genetic ancestries from the exome data using peddy v0.4.2 with the ancestry labelled 1,000 Genomes Project as reference.<sup>55</sup> Of the 287,917 UKB sequences, 18,212 (6.3%) had a Pr(European) ancestry prediction of less than 0.99. Focusing on the remaining 269,706 UKB participants, we further restricted the European ancestry cohort to those within  $\pm 4$  s.d. across the top four principal component means. This resulted in the exclusion of an additional 535 (0.2%) outlier participants. In total, there were 269,171 predominantly unrelated participants of European ancestry who were included in our European case–control analyses. We also used peddy-derived ancestry predictions to perform case–control PheWAS within non-European populations where there were at least 1,000 exome-sequenced individuals available (see the section ‘Collapsing analyses’). Through this step, we identified and used 4,744 ( $\text{Pr}(\text{African}) > 0.95$ ), 1,475 ( $\text{Pr}(\text{East Asian}) > 0.95$ ) and 5,714 ( $\text{Pr}(\text{South Asian}) > 0.95$ ) UKB participants for ancestry-independent collapsing analyses.

## ExWAS analyses

The contribution of rare variants to common disease has, until recently, only been assessed for a subset of complex traits. The gnomAD, which includes exome and genome sequencing data of 141,456 individuals, constitutes the largest publicly available next-generation sequencing resource to date<sup>27</sup>. While this resource has undeniably transformed our ability to interpret rare variants and discover disease-associated genes, it is unsuited to the systematic assessment of the contribution of rare variation to human disease as it lacks linked phenotypic data.

We tested the 2,108,983 variants identified in at least six individuals from the 269,171 predominantly unrelated European ancestry UKB exomes. Variants were required to pass the following quality control criteria: minimum coverage 10X; percent of alternate reads in heterozygous variants  $\geq 0.2$ ; binomial test of alternate allele proportion departure from 50% in heterozygous state  $P > 1 \times 10^{-6}$ ; genotype quality score (GQ)  $\geq 20$ ; Fisher's strand bias score (FS)  $\leq 200$  (indels)  $\leq 60$  (SNVs); mapping quality score (MQ)  $\geq 40$ ; quality score (QUAL)  $\geq 30$ ; read position rank sum score (RPRS)  $\geq -2$ ; mapping quality rank sum score (MQRS)  $\geq -8$ ; DRAGEN variant status = PASS; variant site is not missing (that is, less than 10X coverage) in 10% or more of sequences; the variant did not fail any of the aforementioned quality control in 5% or more of sequences; the variant site achieved tenfold coverage in 30% or more of gnomAD exomes, and if the variant was observed in gnomAD exomes, 50% or more of the time those variant calls passed the gnomAD quality control filters (gnomAD exome AC/AC\_raw  $\geq 50\%$ ).

Variant-level  $P$  values were generated adopting a Fisher's exact two-sided test. Three distinct genetic models were studied for binary traits: allelic (A versus B allele), dominant (AA + AB versus BB) and recessive (AA versus AB + BB), where A denotes the alternative allele and B denotes the reference allele. For quantitative traits, we adopted a linear regression (correcting for age, sex and age  $\times$  sex) and replaced the allelic model with a genotypic (AA versus AB versus BB) test. For ExWAS analysis, we used a significance cut-off of  $P \leq 2 \times 10^{-9}$ . To support the use of this threshold in this study, we performed an  $n$ -of-1 permutation on the binary and quantitative trait dominant model ExWAS. Only 18 of 38.7 billion permuted tests had  $P \leq 2 \times 10^{-9}$ , and 58 of 38.7 billion permuted tests had  $P$  values less than a more liberal cut-off of  $1 \times 10^{-8}$  (Supplementary Tables 18, 19). At this conservative  $P \leq 2 \times 10^{-9}$  threshold, the expected number of ExWAS PheWAS false positives is 18 out of the 46,947 observed significant associations.

## Collapsing analyses

To perform collapsing analyses, we aggregate variants within each gene that fit a given set of criteria, identified as qualifying variants<sup>17</sup>. Overall, we performed 11 non-synonymous collapsing analyses, including 10 dominant and one recessive model, plus

an additional synonymous variant model as an empirical negative control. In each model, for each gene, the proportion of cases was compared to the proportion of controls among individuals carrying one or more qualifying variants in that gene. The exception is the recessive model, where a participant must have two qualifying alleles, either in homozygous or potential compound heterozygous form. Hemizygous genotypes for the X chromosome were also qualified for the recessive model. The qualifying variant criteria for each collapsing analysis model are in Extended Data Table 1. These models were designed to collectively capture a wide range of genetic architectures. They vary in terms of allele frequency (from private up to a maximum of 5%), predicted consequence (for example, PTV or missense), and REVEL and MTR scores. On the basis of SnpEff annotations, we defined synonymous variants as those annotated as ‘synonymous\_variant’. We defined PTVs as variants annotated as exon\_loss\_variant, frameshift\_variant, start\_lost, stop\_gained, stop\_lost, splice\_acceptor\_variant, splice\_donor\_variant, gene\_fusion, bidirectional\_gene\_fusion, rare\_amino\_acid\_variant, and transcript\_ablation. We defined missense as: missense\_variant\_splice\_region\_variant, and missense\_variant. Non-synonymous variants included: exon\_loss\_variant, frameshift\_variant, start\_lost, stop\_gained, stop\_lost, splice\_acceptor\_variant, splice\_donor\_variant, gene\_fusion, bidirectional\_gene\_fusion, rare\_amino\_acid\_variant, transcript\_ablation, conservative\_inframe\_deletion, conservative\_inframe\_insertion, disruptive\_inframe\_insertion, disruptive\_inframe\_deletion, missense\_variant\_splice\_region\_variant, missense\_variant, and protein\_altering\_variant.

Collapsing analysis *P* values were generated by using a Fisher’s exact two-sided test. For quantitative traits, we used a linear regression, correcting for age, sex and age × sex.

For all models (Extended Data Table 1), we applied the following quality control filters: minimum coverage 10X; annotation in CCDS transcripts (release 22; approximately 34 Mb); at most 80% alternate reads in homozygous genotypes; percent of alternate reads in heterozygous variants  $\geq 0.25$  and  $\leq 0.8$ ; binomial test of alternate allele proportion departure from 50% in heterozygous state  $P > 1 \times 10^{-6}$ ; GQ  $\geq 20$ ; FS  $\leq 200$  (indels)  $\leq 60$  (SNVs); MQ  $\geq 40$ ; QUAL  $\geq 30$ ; read position rank sum score  $\geq -2$ ; MQRS  $\geq -8$ ; DRAGEN variant status = PASS; the variant site achieved tenfold coverage in  $\geq 25\%$  of gnomAD exomes, and if the variant was observed in gnomAD exomes, the variant achieved exome z-score  $\geq -2.0$  and exome MQ  $\geq 30$ .

To quantify how well a protein-coding gene is represented across all individuals by the exome sequence data, we estimated informativeness statistics for each studied gene on the basis of sequencing coverage across the available exomes (Supplementary Methods, Supplementary Table 24). Moreover, we created dummy phenotypes to correspond to each of the four exome sequence delivery batches to identify and

exclude from analyses genes and variants that reflected sequencing batch effects; we provide these as a cautionary list resource for other UKB exome researchers (Supplementary Methods, Supplementary Tables [25–27](#)).

For the pan-ancestry analysis, a Cochran–Mantel–Haenszel test was performed to generate a combined  $2 \times 2 \times N$  stratified  $P$  value, with  $N$  representing up to all four genetic ancestry groups. This was performed for 4,836 binary phenotypes where one of the three non-European ancestries had five or more cases and for all quantitative traits. For the quantitative traits, we used a linear regression model that included the following covariates: categorical ancestry (European, African, East Asian or South Asian), the top five ancestry principal components, age and sex.

## Compute processing times

Our end-to-end (CRAM → FASTQ → BAM → VCF) processing of the 302,355 UKB exomes was achieved at an average rate of 1,600 exomes per hour, consuming a total of 52,000 hours of CPU time running on Linux servers with FPGA acceleration.

Regarding our collapsing PheWAS analyses, construction of the full set of genotype and phenotype matrices took 13,000 and 30 CPU hours to compile, respectively. The preprocessing steps such as rebalancing sex-specific case–control ratios are incorporated in the phenotype matrix construction time. Subsequently, the approximately 4.5 billion collapsing analysis statistical tests were calculated in 19,000 CPU hours. In wall-clock hours, this took 30 h to generate all the collapsing and phenotype matrices. Once the intermediate files were ready, the roughly 4.5 billion collapsing statistical tests took 8 h to complete.

Regarding our variant-level ExWAS, upon construction of our variant matrices, which took 2,500 CPU hours to compile, all 108 billion statistical tests were calculated in 855,000 CPU hours. In wall-clock hours, this took 37 h to generate the variant matrices. Once these intermediate files were ready, the approximately 108 billion ExWAS statistical tests took 27 h for binary traits and 11 h for quantitative traits.

## Defining the study-wide significant cut-offs for collapsing analyses

Bonferroni correction for multiple testing was inappropriate to use in this study given the high degree of correlation among the studied phenotypes and the level of similarity among the multiple collapsing models. Thus, we took two approaches to define more appropriate study-wide significance thresholds for the gene-based collapsing PheWAS.

We used a synonymous collapsing analysis model as an empirical negative control. Here it is expected that synonymous variants will generally not significantly contribute to disease risk and could thus act as a useful empirical negative control for study-wide

*P* value thresholding. Across the 17,361 studied binary phenotypes and 18,762 studied genes, we observed a distribution of 325,727,082 Fisher's exact test statistics corresponding to the synonymous collapsing model. At the tail of this distribution for binary traits, we identified two genuine relationships: *IGLL5* synonymous variants enriched among 'Union#C911#C91.1 chronic lymphocytic leukaemia' ( $P = 2.5 \times 10^{-11}$ ) and its parent node 'Union#C91#C91 lymphoid leukaemia' ( $P = 1.2 \times 10^{-10}$ ). Following this, we observed a tail of *P* values beginning from  $P = 2.2 \times 10^{-8}$  (Supplementary Table 20). Similarly, for the 1,419 quantitative phenotypes, we observed a distribution of 26,623,278 Fisher's exact test statistics corresponding to the synonymous collapsing model. At the tail of this distribution, we identified two genuine relationships: *MACROD1* synonymous variants correlating with decreased levels of 'Urate' ( $P = 2.8 \times 10^{-30}$ <sup>56</sup>) and *ALPL* synonymous variants correlating with decreased levels of 'alkaline phosphatase' ( $P = 9.3 \times 10^{-9}$ <sup>57</sup>). Following this, we saw a tail of *P* values beginning from  $P = 5.2 \times 10^{-8}$  (Supplementary Table 20).

With this magnitude of test statistics generated in the PheWAS scale, another proposal for *P* value thresholding involves *n*-of-1 permutation<sup>58</sup>. In applying this approach, we shuffled the case-control (or quantitative measurement) labels once for every phenotype while maintaining the participant-genotype structure and across all 11 non-synonymous collapsing models for binary traits (3,582,997,902 tests) and quantitative traits (292,856,058 tests). Reviewing the tails of these two *P* value distributions, the lowest permutation-based *P* value achieved was  $1.9 \times 10^{-9}$  (binary tests) and  $3.2 \times 10^{-9}$  (quantitative tests).

Given the scale and correlations among this dataset, we found that both of these approaches provide suitable alternatives to the Bonferroni *P* value threshold, which in this case would be  $P < 1.2 \times 10^{-11}$ . Prioritizing the results of the permutation-based approach because it captures the data structure across all our models, we define a conservative study-wide significance cut-off of  $P \leq 2 \times 10^{-9}$  for the non-synonymous collapsing analysis results presented in this paper (Supplementary Tables 20, 21). Under this conservative threshold, no positive associations are expected under the null for collapsing analyses.

Finally, for each of the 225,360 exome-wide collapsing analyses comprising the collapsing PheWAS (12 models  $\times$  (17,361 + 1,419) studied phenotypes), we calculated the lambda genomic inflation factor ( $\lambda$ ) after excluding genes achieving exome-wide significance  $P < 2.6 \times 10^{-6}$  for that phenotype (Supplementary Tables 22, 23).

## Collapsing analysis enrichment for approved drug targets

We tested for the enrichment of drug targets among collapsing analysis associations using five publicly available lists: a custom list ( $n = 387$ ;

[https://raw.githubusercontent.com/ericminikel/drug\\_target\\_lof/master/data/drugbank/drug\\_gene\\_match.tsv](https://raw.githubusercontent.com/ericminikel/drug_target_lof/master/data/drugbank/drug_gene_match.tsv)) that was originally derived from DrugBank<sup>59</sup>, and another four lists<sup>6</sup> that were originally derived from the Informa Pharmaprojects database<sup>46</sup>. These four lists included drug targets from their latest stages of clinical trials, labelled as ‘Approved’ ( $n = 2,620$ ), ‘Phase I Clinical Trial’ ( $n = 3,365$ ), ‘Phase II Clinical Trial’ ( $n = 5,479$ ) and ‘Phase III Clinical Trial’ ( $n = 1,233$ ).

For each gene tested in the collapsing analysis, we only retained the most significantly associated phenotype. Distinct gene–phenotype relationships from the collapsing analysis were partitioned into three categories (significant:  $P < 2 \times 10^{-9}$  (binary  $n = 82$ , quantitative  $n = 269$ ); suggestive:  $2 \times 10^{-9} < P < 1 \times 10^{-7}$  (binary  $n = 113$ , quantitative  $n = 126$ ); or non-significant:  $P > 1 \times 10^{-7}$  (binary  $n = 18,551$ , quantitative  $n = 18,351$ )). The relationship between drug target status and gene–phenotype significance was assessed using Fisher’s exact test for each gene list. Specifically, for each of the five lists, we created a contingency table that included the number of significant collapsing analysis genes that intersected with the list and the number of genes that did not intersect with the list out of the list of genes tested in the PheWAS ( $n = 18,762$ ). This was performed for both binary and quantitative traits. We also performed enrichment testing for OMIM<sup>32</sup> genes and GWAS Catalog<sup>31</sup> significant hits (both last accessed on 14 July 2020). We included the most significant associations per gene for the GWAS analysis.

## Ethics reporting

The protocols for UKB are overseen by The UK Biobank Ethics Advisory Committee (EAC); for more information see <https://www.ukbiobank.ac.uk/ethics/> and <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Association statistics generated in this study are publicly available through our AstraZeneca Centre for Genomics Research (CGR) PheWAS Portal (<http://azphewas.com/>). All whole-exome sequencing data described in this paper are publicly available to registered researchers through the UKB data access protocol. Exomes can be found in the UKB showcase portal: <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=170>. Additional information

about registration for access to the data is available at <http://www.ukbiobank.ac.uk/register-apply/>. Data for this study were obtained under Resource Application Number 26041.

A custom list of drug targets from DrugBank is available: [https://raw.githubusercontent.com/ericminikel/drug\\_target\\_lof/master/data/drugbank/drug\\_gene\\_match.tsv](https://raw.githubusercontent.com/ericminikel/drug_target_lof/master/data/drugbank/drug_gene_match.tsv). A Pharmaprojects-based list of drug targets is available: [https://raw.githubusercontent.com/AbbVie-ComputationalGenomics/genetic-evidence-approval/master/data/target\\_indication.tsv](https://raw.githubusercontent.com/AbbVie-ComputationalGenomics/genetic-evidence-approval/master/data/target_indication.tsv).

We used data from the OMIM (<https://www.omim.org>)<sup>32</sup>, MTR (<http://mtr-viewer.mdhs.unimelb.edu.au>)<sup>35</sup>, REVEL<sup>50</sup>, gnomAD (<https://gnomad.broadinstitute.org>)<sup>27</sup>, EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas>)<sup>31</sup>, ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>)<sup>30</sup> and FinnGen release r5 (<https://www.finngen.fi/en>).

## Code availability

PheWAS and ExWAS association tests were performed using a custom framework, PEACOK (PEACOK 1.0.7), which is an extension and enhancement of PHE-SANT. PEACOK 1.0.7 is available on GitHub: <https://github.com/astrazeneca-cgr-publications/PEACOK/>.

## References

1. 1. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
2. 2. Szustakowski, J. D. et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* **53**, 942–948 (2021).
3. 3. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
4. 4.

Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17**, 507–522 (2016).

5. 5.

Harper, A. R., Nayee, S. & Topol, E. J. Protective alleles and modifier variants in human health and disease. *Nat. Rev. Genet.* **16**, 689–701 (2015).

6. 6.

King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).

7. 7.

Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

8. 8.

Claussnitzer, M. et al. A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).

9. 9.

Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).

10. 10.

Cirulli, E. T. et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* **11**, 542 (2020).

11. 11.

Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

12. 12.

Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341

(2018).

13. 13.

Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).

14. 14.

Kosmicki, J. A., Churchhouse, C. L., Rivas, M. A. & Neale, B. M. Discovery of rare variants for complex phenotypes. *Hum. Genet.* **135**, 625–634 (2016).

15. 15.

Greene, D., Richardson, S. & Turro, E. A fast association test for identifying pathogenic variants involved in rare diseases. *Am. J. Hum. Genet.* **101**, 104–114 (2017).

16. 16.

Povysil, G. et al. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat. Rev. Genet.* **20**, 747–759 (2019).

17. 17.

Petrovski, S. et al. An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **196**, 82–93 (2017).

18. 18.

Cameron-Christie, S. et al. Exome-based rare-variant analyses in CKD. *J. Am. Soc. Nephrol.* **30**, 1109–1122 (2019).

19. 19.

Cirulli, E. T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441 (2015).

20. 20.

Epi4K Consortium & Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case–control sequencing study. *Lancet Neurol.* **16**, 135–143 (2017).

21. 21.

Carss, K. J. et al. Spontaneous coronary artery dissection: insights on rare genetic variation from genome sequencing. *Circ. Genom. Precis. Med.* **13**, e003030 (2020).

22. 22.

Povysil, G. et al. Assessing the role of rare genetic variation in patients with heart failure. *JAMA Cardiol.* **6**, 379–386 (2021).

23. 23.

Dhindsa, R. S. et al. Identification of a missense variant in SPDL1 associated with idiopathic pulmonary fibrosis. *Commun. Biol.* **4**, 392 (2021).

24. 24.

Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software application profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).

25. 25.

Petrovski, S. & Goldstein, D. B. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* **17**, 157 (2016).

26. 26.

DeBoever, C. et al. Medical relevance of protein-truncating variants across 337,205 individuals in the UK Biobank study. *Nat. Commun.* **9**, 1612 (2018).

27. 27.

Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

28. 28.

Emdin, C. A. et al. Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat. Commun.* **9**, 1613 (2018).

29. 29.

MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

30. 30.

Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

31. 31.

Bunielo, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

32. 32.

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).

33. 33.

Li, X. et al. The impact of rare variation on gene expression across tissues. *Nature* **550**, 239–243 (2017).

34. 34.

Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

35. 35.

Traynelis, J. et al. Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. *Genome Res.* **27**, 1715–1729 (2017).

36. 36.

Weidinger, S. et al. Filaggrin mutations, atopic eczema, hay fever, and asthma in children. *J. Allergy Clin. Immunol.* **121**, 1203–1209.e1 (2008).

37. 37.

Kezic, S. Loss-of-function mutations in filaggrin gene and malignant melanoma. *J. Eur. Acad. Dermatol. Venereol.* **32**, 193 (2018).

38. 38.

Kaae, J. et al. Filaggrin gene mutations and risk of basal cell carcinoma. *Br. J. Dermatol.* **169**, 1162–1164 (2013).

39. 39.

Thyssen, J. P. & Elias, P. M. It remains unknown whether filaggrin gene mutations evolved to increase cutaneous synthesis of vitamin D. *Genome Biol. Evol.* **9**, 900–901 (2017).

40. 40.

Pagnamenta, A. T. et al. Exome sequencing can detect pathogenic mosaic mutations present at low allele frequencies. *J. Hum. Genet.* **57**, 70–72 (2012).

41. 41.

Ben-Eghan, C. et al. Don't ignore genetic data from minority populations. *Nature* **585**, 184–186 (2020).

42. 42.

Diogo, D. et al. Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.* **9**, 4285 (2018).

43. 43.

Cameron-Christie, S. et al. A broad exome study of the genetic architecture of asthma reveals novel patient subgroups. Preprint at <https://doi.org/10.1101/2020.12.10.419663> (2020).

44. 44.

Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).

45. 45.

Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

46. 46.

Informa. Pharmaprojects: track pharma R&D. *Informa* <https://pharmaintelligence.informa.com/products-and-services/data-and-analysis/pharmaprojects> (2021).

47. 47.

Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

48. 48.

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 (Austin)* **6**, 80–92 (2012).

49. 49.

Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).

50. 50.

Ioannidis, N. M. et al. REVEL: an Ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

51. 51.

Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).

52. 52.

Pujar, S. et al. Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding regions supported by expert curation. *Nucleic Acids Res.* **46**, D221–D228 (2018).

53. 53.

Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

54. 54.

Hanscombe, K. B., Coleman, J. R. I., Traylor, M. & Lewis, C. M. ukbtools: An R package to manage and query UK Biobank data. *PLoS ONE* **14**, e0214311 (2019).

55. 55.

Pedersen, B. S. & Quinlan, A. R. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. *Am. J. Hum. Genet.* **100**, 406–413 (2017).

56. 56.

Tin, A. et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* **51**, 1459–1474 (2019).

57. 57.

Olafsson, S. et al. Common and rare sequence variants influencing tumor biomarkers in blood. *Cancer Epidemiol. Biomarkers Prev.* **29**, 225–235 (2020).

58. 58.

Annis, A. et al. Determining genome-wide significance thresholds in biobanks with thousands of phenotypes: a case study using the Michigan Genomics Initiative. Presented at Annual Meeting of The American Society of Human Genetics 2019 (2019).

59. 59.

Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

## Acknowledgements

We thank the participants and investigators in the UKB study who made this work possible (Resource Application Number 26041); the UKB Exome Sequencing Consortium (UKB-ESC) members AbbVie, Alnylam Pharmaceuticals, AstraZeneca, Biogen, Bristol-Myers Squibb, Pfizer, Regeneron and Takeda for funding the generation of the data and Regeneron Genetics Center for completing the sequencing and initial quality control of the exome sequencing data; the AstraZeneca Centre for Genomics Research Analytics and Informatics team for processing and analysis of

sequencing data; and M. Hurles and D. Balding for feedback on this manuscript. We acknowledge the participants and investigators of the FinnGen study.

## Author information

### Author notes

1. These authors contributed equally: Quanli Wang, Ryan S. Dhindsa, Keren Carss

### Affiliations

1. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Waltham, MA, USA

Quanli Wang & Ryan S. Dhindsa

2. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Keren Carss, Andrew R. Harper, Abhishek Nag, Ioanna Tachmazidou, Dimitrios Vitsios, Sri V. V. Deevi, Ronen Artzi, Oliver Burren, Sophia Cameron-Christie, Zara Ghazoui, Fengyuan Hu, Magda Jeznach, Xiao Jiang, Samuel H. Lewis, Kieren Lythgow, Peter Maccallum, Athena Matakidou, Sean O'Dell, Joel Okae, Amanda O'Neill, Dirk S. Paul, Sebastian Wasilewski, Katherine R. Smith, Carolina Haefliger & Slavé Petrovski

3. Translational Science and Experimental Medicine, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Alex Mackay, Daniel Muthas, Michael Hühn, Susan Monkley, Henric Olsson, Bastian R. Angermann & Benjamin Georgi

4. Precision Medicine & Biosamples, Oncology R&D, AstraZeneca, Cambridge, UK

Ronen Artzi & Ruth March

5. Translational Science and Experimental Medicine, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Glenda Lassi & Adam Platt

6. Department of Medicine, University of Melbourne, Austin Health, Melbourne, Victoria, Australia

Slavé Petrovski

7. Epilepsy Research Centre, University of Melbourne, Austin Health, Melbourne, Victoria, Australia

Slavé Petrovski

8. Translational Medicine, Research and Early Development, Oncology R&D, AstraZeneca, Waltham, MA, USA

Carl Barrett, Brian Dougherty, Zhongwu Lai, Bolan Linghu & Jorge Zeron

9. Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Maria Belvisi

10. Translational Genomics, Discovery Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Mohammad Bohlooly-Y

11. Biosciences CKD, Research and Early Development, Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Lisa Buvall & Pernille B. L. Hansen

12. Translational Science & Experimental Medicine, Research and Early Development, Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Benjamin Challis, Chanchal Kumar, Sven Moosmang, Anna Reznichenko & Anna Walentinsson

13. Bioscience Asthma, Research and Early Development, Respiratory & Immunology, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK

Suzanne Cohen

14. Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Andrew Davis & Michael A Snowden

15. Research and Early Development, Cardiovascular, Renal and Metabolism,  
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Regina F. Danielson

16. Oncology Discovery, Early Oncology, Oncology R&D, AstraZeneca, Cambridge,  
UK

Carla Martins

17. Early Clinical Development, Research and Early Development, Cardiovascular,  
Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg,  
Sweden

Erik Michaëlsson

18. Bioscience Asthma, Research and Early Development, Respiratory &  
Immunology, Biopharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, USA

Yoichiro Ohne

19. Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK

Menelas N. Pangalos

## **Consortia**

### **AstraZeneca Genomics Initiative**

- Bastian R. Angermann
- , Ronen Artzi
- , Carl Barrett
- , Maria Belvisi
- , Mohammad Bohlooly-Y
- , Oliver Burren
- , Lissa Buvall
- , Benjamin Challis
- , Sophia Cameron-Christie
- , Suzanne Cohen
- , Andrew Davis
- , Regina F. Danielson

- , Brian Dougherty
- , Benjamin Georgi
- , Zara Ghazoui
- , Pernille B. L. Hansen
- , Fengyuan Hu
- , Magda Jeznach
- , Xiao Jiang
- , Chanchal Kumar
- , Zhongwu Lai
- , Glenda Lassi
- , Samuel H. Lewis
- , Bolan Linghu
- , Kieren Lythgow
- , Peter Maccallum
- , Carla Martins
- , Athena Matakidou
- , Erik Michaëlsson
- , Sven Moosmang
- , Sean O'Dell
- , Yoichiro Ohne
- , Joel Okae
- , Amanda O'Neill
- , Dirk S. Paul
- , Anna Reznichenko
- , Michael A Snowden
- , Anna Walentinsson
- , Jorge Zeron
- & Menelas N. Pangalos

## **Contributions**

Q.W., R.S.D. and S.P. designed the study. Q.W., R.S.D., K.C., A.R.H., A.N., I.T., D.V., M.H., S.M., K.R.S. and S.P. performed analyses and statistical interpretation. Q.W., S.V.V.D. and S.W. did the bioinformatics processing. I.T. performed benchmarking with support from Q.W. Q.W., K.C., K.R.S. and S.W. scoped and lead the PheWAS portal development. R.M., A.P., C.H. and S.P. contributed to the organization of the project. Q.W., R.S.D., K.C., A.R.H., A.N., I.T. and S.P. wrote the manuscript. Q.W., R.S.D., K.C., A.R.H., A.N., I.T., D.V., S.V.V.D., A.M., D.M., M.H., S.M., H.O., S.W., K.R.S., R.M., A.P., C.H. and S.P. reviewed the manuscript.

## **Corresponding author**

Correspondence to Slavé Petrovski.

## Ethics declarations

### Competing interests

Q.W., R.S.D., K.C., A.R.H., A.N., I.T., D.V., S.V.V.D., A.M., D.M., M.H., S.M., H.O., S.W., K.R.S., R.M., A.P., C.H. and S.P are current employees and/or stockholders of AstraZeneca.

## Additional information

**Peer review information** *Nature* thanks Beryl Cummings, Dean Sheppard, David van Heel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Phenotypic and demographic diversity of the sequenced UK Biobank cohort.](#)

**a**, The percentage of binary union traits assessed in the cohort per disease chapter. **b**, The percentage of quantitative traits assessed in the cohort per chapter. **c**, The median number of cases of European ancestry per binary union phenotype stratified by chapter with interquartile range depicted. The median number of European cases per binary union phenotype was 191 (interquartile range: 72-773). **d**, The median number of participants of European ancestry tested for quantitative traits stratified by chapter with interquartile ranges depicted. The median number of individuals tested for quantitative traits was 13,782 (interquartile range: 13,780-17,795). **e**, Histogram depicting the number of binary union phenotypes per patient. The x-axis was capped at 200 for visual clarity. The median number of binary union traits per European participant was 25 (interquartile range: 12-45) of a possible 4,911. **f**, The distribution of represented genetic ancestries in the sequenced cohort. EUR = European, SAS = South Asian, AFR = African, EAS = East Asian, AMR = American. **g**, The distribution of the number of rare ( $MAF < 0.005\%$ ) qualifying variants (QVs) in OMIM-derived Mendelian disease genes per ancestral group. Error bars in **(c, d)** represent the interquartile range.

## Extended Data Fig. 2 Rare PTVs and direction of variant effects.

**a**, The number of genes (y-axis) with at least  $N$  rare (MAF >0.01) protein-truncating variant (PTV) carriers (x-axis) in the cohort. Colours correspond to heterozygous (Het), putative compound heterozygous plus homozygous/hemizygous carriers (comp. het), and exclusively homozygous/hemizygous carriers (recessive). **b**, Distribution of the directions of effect for rare (MAF <0.1%) non-synonymous variant associations with quantitative phenotypes. Only phenotypes with at least five significant non-synonymous variant associations ( $P \leq 2 \times 10^{-9}$ ) in a given gene were considered.

## Extended Data Fig. 3 Quantitative trait collapsing analysis.

Plot depicting significant gene-phenotype associations for quantitative traits. For gene-phenotype associations that appear in multiple collapsing models, we display only the association with the strongest effect size. The dashed line represents the genome-wide significant p-value threshold ( $2 \times 10^{-9}$ ). The plot is capped at  $-\log_{10}(P) = 50$  and only associations with  $P < 10^{-5}$  are included ( $n = 22,549$ ).

## Extended Data Fig. 4 Drug target enrichments.

Forest plots demonstrating enrichment of drug targets curated in DrugBank and the Informa Pharmaprojects databases among significant (Tier 1) and nearly significant (Tier 2) binary trait associations, quantitative trait associations, OMIM genes, and GWAS signals. P-values were calculated via Fisher's exact test (two-sided). Error bars represent 95% confidence intervals of the Odds Ratio. The total numbers of genes per category are as follows: DrugBank-derived ( $n = 386$ ); Approved from Informa Pharmaprojects ( $n = 463$ ); Phase III from Informa Pharmaprojects ( $n = 474$ ); Phase II from Informa Pharmaprojects ( $n = 1006$ ); Phase I from Informa Pharmaprojects ( $n = 921$ ); Collapsing – Binary (Tier 1  $n = 82$ ; Tier 2  $n = 113$ ); Collapsing - Quantitative (Tier 1  $n = 269$ ; Tier 2  $n = 126$ ); OMIM ( $n = 3875$ ); GWAS (Tier 1  $n = 8975$ ; Tier 2  $n = 1717$ ).

## Extended Data Fig. 5 Collapsing analysis comparisons.

**a**, Distribution of lambda (inflation factor) values across all collapsing models for binary and quantitative traits. **b**, Venn diagram for gene-trait associations identified by three studies using the first tranche of 50K UKB. There are 81 distinct significant gene-trait associations ( $P < 3.4 \times 10^{-10}$ ) found among phenotypes that were studied by the three efforts (Supplementary Table 28). **c**, Percentage of suggestive binary gene-phenotype associations that became significant (sig) ( $P < 2 \times 10^{-9}$ ), non-significant (non-sig) ( $P > 1 \times 10^{-7}$ ) or remained suggestive (sugg) ( $2 \times 10^{-9} < P < 1 \times 10^{-7}$ ) with each

successive UKB tranche release for binary traits (supplementary methods). 300Kv1 includes phenotypic data released up to April 2017, and 300Kv2 includes additional phenotypic data for the same set of samples released up to July 2020.

## **Extended Data Fig. 6 Pan-ancestry delta Phred distributions.**

**a, b**, Distribution of the change between Phred ( $(-10 \cdot \log_{10}[\text{p-values}])$ ) scores from the pan-ancestry collapsing analysis and the European-only collapsing analysis for binary traits (**a**) and quantitative traits (**b**). The x-axis in both figures are capped at -50 and +50.

## **Extended Data Table 1 Collapsing analysis models**

## **Supplementary information**

### **Supplementary Information**

This file contains Supplementary Methods, Supplementary Figures 1-4, Supplementary Tables 3, 6, 9, 11, 13, 18, 21, 23, 25, and 27, detailed descriptions of Supplementary Datasets, and Supplementary References.

### **Reporting Summary**

### **Supplementary Tables**

This file contains Supplementary Tables 1, 2, 4, 5, 7, 8, 10, 12, 14, 15-17, 19, 20, 22, 24, 26, 28 and 29.

### **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

Wang, Q., Dhindsa, R.S., Carss, K. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).  
<https://doi.org/10.1038/s41586-021-03855-y>

- Received: 03 November 2020
- Accepted: 28 July 2021
- Published: 10 August 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03855-y>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03855-y>.

- Article
- [Published: 08 September 2021](#)

# Bioaccumulation of therapeutic drugs by human gut bacteria

- [Martina Klünemann](#) ORCID: [orcid.org/0000-0002-1602-5371](https://orcid.org/0000-0002-1602-5371)<sup>1 nAff9</sup>,
- [Sergej Andrejev](#) ORCID: [orcid.org/0000-0002-7875-0261](https://orcid.org/0000-0002-7875-0261)<sup>1 nAff10</sup>,
- [Sonja Blasche](#)<sup>1,2 nAff1</sup>,
- [Andre Mateus](#) ORCID: [orcid.org/0000-0001-6870-0677](https://orcid.org/0000-0001-6870-0677)<sup>1 nAff1</sup>,
- [Prasad Phapale](#) ORCID: [orcid.org/0000-0002-9487-597X](https://orcid.org/0000-0002-9487-597X)<sup>1</sup>,
- [Saravanan Devendran](#)<sup>1</sup>,
- [Johanna Vappiani](#)<sup>3</sup>,
- [Bernd Simon](#)<sup>1</sup>,
- [Timothy A. Scott](#)<sup>4</sup>,
- [Eleni Kafkia](#) ORCID: [orcid.org/0000-0001-9550-4487](https://orcid.org/0000-0001-9550-4487)<sup>2</sup>,
- [Dimitrios Konstantinidis](#)<sup>1</sup>,
- [Katharina Zirngibl](#)<sup>1,2</sup>,
- [Eleonora Mastrorilli](#) ORCID: [orcid.org/0000-0003-2127-4150](https://orcid.org/0000-0003-2127-4150)<sup>1</sup>,
- [Manuel Banzhaf](#)<sup>1 nAff11</sup>,
- [Marie-Therese Mackmull](#) ORCID: [orcid.org/0000-0003-2928-1144](https://orcid.org/0000-0003-2928-1144)<sup>1 nAff12</sup>,
- [Felix Hövelmann](#)<sup>1</sup>,
- [Leo Nesme](#)<sup>1 nAff13</sup>,
- [Ana Rita Brochado](#)<sup>1 nAff14</sup>,
- [Lisa Maier](#) ORCID: [orcid.org/0000-0002-6473-4762](https://orcid.org/0000-0002-6473-4762)<sup>1 nAff15</sup>,
- [Thomas Bock](#) ORCID: [orcid.org/0000-0002-9314-5318](https://orcid.org/0000-0002-9314-5318)<sup>1 nAff16</sup>,
- [Vinita Periwal](#)<sup>1,2</sup>,
- [Manjeet Kumar](#)<sup>1</sup>,
- [Yongkyu Kim](#) ORCID: [orcid.org/0000-0002-3336-6741](https://orcid.org/0000-0002-3336-6741)<sup>1</sup>,

- [Melanie Tramontano](#) ORCID: [orcid.org/0000-0001-6407-527X](#)<sup>1 nAff10</sup>,
- [Carsten Schultz](#)<sup>1 nAff17</sup>,
- [Martin Beck](#) ORCID: [orcid.org/0000-0002-7397-1321](#)<sup>1 nAff18</sup>,
- [Janosch Hennig](#) ORCID: [orcid.org/0000-0001-5214-7002](#)<sup>1 nAff19</sup>,
- [Michael Zimmermann](#) ORCID: [orcid.org/0000-0002-5797-3589](#)<sup>1</sup>,
- [Daniel C. Sévin](#)<sup>3</sup>,
- [Filipe Cabreiro](#) ORCID: [orcid.org/0000-0002-3696-4843](#)<sup>4,5 nAff20</sup>,
- [Mikhail M. Savitski](#) ORCID: [orcid.org/0000-0003-2011-9247](#)<sup>1</sup>,
- [Peer Bork](#) ORCID: [orcid.org/0000-0002-2627-833X](#)<sup>1,6,7,8</sup>,
- [Athanasios Typas](#) ORCID: [orcid.org/0000-0002-0797-9018](#)<sup>1</sup> &
- [Kiran R. Patil](#) ORCID: [orcid.org/0000-0002-6166-8640](#)<sup>1,2</sup>

[Nature](#) volume **597**, pages 533–538 (2021)

- 14k Accesses
- 547 Altmetric
- [Metrics details](#)

## Subjects

- [Biochemical networks](#)
- [Microbiome](#)

## Abstract

Bacteria in the gut can modulate the availability and efficacy of therapeutic drugs. However, the systematic mapping of the interactions between drugs and bacteria has only started recently<sup>1</sup> and the main underlying mechanism proposed is the chemical transformation of drugs by microorganisms (biotransformation). Here we investigated the depletion of 15 structurally diverse drugs by 25 representative strains of gut bacteria. This revealed 70 bacteria–drug interactions, 29 of which had not to our knowledge been reported before. Over half of the new interactions can be ascribed to

bioaccumulation; that is, bacteria storing the drug intracellularly without chemically modifying it, and in most cases without the growth of the bacteria being affected. As a case in point, we studied the molecular basis of bioaccumulation of the widely used antidepressant duloxetine by using click chemistry, thermal proteome profiling and metabolomics. We find that duloxetine binds to several metabolic enzymes and changes the metabolite secretion of the respective bacteria. When tested in a defined microbial community of accumulators and non-accumulators, duloxetine markedly altered the composition of the community through metabolic cross-feeding. We further validated our findings in an animal model, showing that bioaccumulating bacteria attenuate the behavioural response of *Caenorhabditis elegans* to duloxetine. Together, our results show that bioaccumulation by gut bacteria may be a common mechanism that alters drug availability and bacterial metabolism, with implications for microbiota composition, pharmacokinetics, side effects and drug responses, probably in an individual manner.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

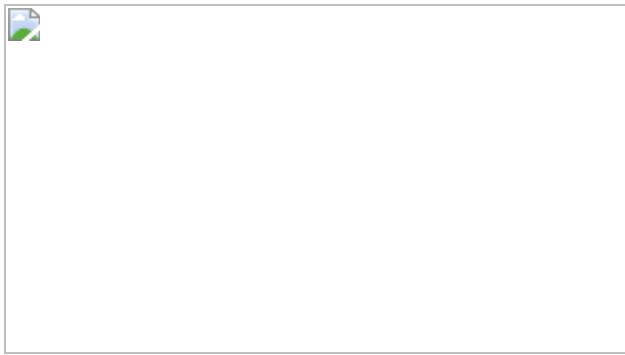
[Rent or Buy](#)

All prices are NET prices.

### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

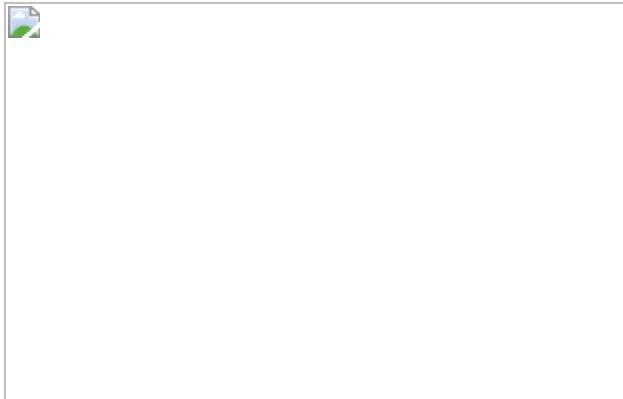
**Fig. 1: Gut bacteria accumulate therapeutic drugs without altering them.**



**Fig. 2: Bioaccumulation of duloxetine affects bacterial physiology.**



**Fig. 3: Duloxetine bioaccumulation alters community assembly and host response.**



## Data availability

All data generated during this study are included in this published Article (and its Supplementary Information files). Supplementary Table 18 provides an overview of the different methods and data associated with all figures. UPLC and mass spectrometry data are deposited at the MetaboLights repository under the accession codes [MTBLS1264](#), [MTBLS1757](#), [MTBLS1627](#), [MTBLS1319](#), [MTBLS1791](#), [MTBLS1792](#), and [MTBLS2885](#). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium with the dataset identifiers [PXD016062](#) and [PXD016064](#). [Source data](#) are provided with this paper.

## Code availability

The data analysis codes are available at  
[https://github.com/sandrejev/drugs\\_bioaccumulation](https://github.com/sandrejev/drugs_bioaccumulation).

## References

1. 1.

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467 (2019).

2. 2.

Forslund, K., Hildebrand, F., Nielsen, T. & Falony, G. A. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262–266 (2015).

3. 3.

Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).

4. 4.

Maier, L. & Typas, A. Systematically investigating the impact of medication on the gut microbiome. *Curr. Opin. Microbiol.* **39**, 128–135 (2017).

5. 5.

Jackson, M. A. et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat. Commun.* **9**, 2655 (2018).

6. 6.

Maier, L. et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).

7. 7.

Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016).

8. 8.

Alexander, J. L. et al. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).

9. 9.

Fuller, A. T. Is p-aminobenzenesulphonamide the active agent in prontosil therapy? *Lancet* **229**, 194–198 (1937).

10. 10.

Goldman, P., Peppercorn, M. A. & Goldin, B. R. Metabolism of drugs by microorganisms in the intestine. *Am. J. Clin. Nutr.* **27**, 1348–1355 (1974).

11. 11.

Chhabra, R. S. Intestinal absorption and metabolism of xenobiotics. *Environ. Health Perspect.* **33**, 61–69 (1979).

12. 12.

Koppel, N., Maini Rekdal, V. & Balskus, E. P. Chemical transformation of xenobiotics by the human gut microbiota. *Science* **356**, eaag2770 (2017).

13. 13.

Sousa, T. et al. The gastrointestinal microbiota as a site for the biotransformation of drugs. *Int. J. Pharm.* **363**, 1–25 (2008).

14. 14.

Klaassen, C. D. & Cui, J. Y. Review: mechanisms of how the intestinal microbiota alters the effects of drugs and bile acids. *Drug Metab. Dispos.* **43**, 1505–1521 (2015).

15. 15.

Haiser, H. J., Seim, K. L., Balskus, E. P. & Turnbaugh, P. J. Mechanistic insight into digoxin inactivation by *Eggerthella lenta* augments our understanding of its pharmacokinetics. *Gut Microbes* **5**, 233–238 (2014).

16. 16.

Koppel, N., Bisanz, J. E., Pandelia, M.-E., Turnbaugh, P. J. & Balskus, E. P. Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *eLife* **7**, e33953 (2018).

17. 17.

Wallace, B. D., Wang, H., Lane, K. T. & Scott, J. E. Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).

18. 18.

Tramontano, M. et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* **3**, 514–522 (2018).

19. 19.

Chrystal, E. J. T., Koch, R. L., McLafferty, M. A. & Goldman, P. Relationship between metronidazole metabolism and bactericidal activity. *Antimicrob. Agents Chemother.* **18**, 566–573 (1980).

20. 20.

Mahmood, S., Khalid, A., Arshad, M., Mahmood, T. & Crowley, D. E. Detoxification of azo dyes by bacterial oxidoreductase enzymes. *Crit. Rev. Biotechnol.* **36**, 639–651 (2016).

21. 21.

Khan, A. K. A., Guthrie, G., Johnston, H. H., Truelove, S. C. & Williamson, D. H. Tissue and bacterial splitting of sulphasalazine. *Clin. Sci.* **64**, 349–354 (1983).

22. 22.

Goodman, A. L. et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.

*Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).

23. 23.

Shu, Y. Z. & Kingston, D. G. A. Metabolism of levamisole, an anti-colon cancer drug, by human intestinal bacteria. *Xenobiotica* **21**, 737–750 (1991).

24. 24.

Schloissnig, S. et al. Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45–50 (2013).

25. 25.

Fenner, K., Canonica, S., Wackett, L. P. & Elsner, M. Evaluating pesticide degradation in the environment: blind spots and emerging opportunities. *Science* **341**, 752–758 (2013).

26. 26.

Gulde, R., Anliker, S., Kohler, H. E. & Fenner, K. Ion trapping of amines in protozoa: a novel removal mechanism for micropollutants in activated sludge. *Environ. Sci. Technol.* **52**, 52–60 (2018).

27. 27.

Congeevaram, S., Dhanarani, S., Park, J., Dexilin, M. & Thamaraiselvi, K. Biosorption of chromium and nickel by heavy metal resistant fungal and bacterial isolates. *J. Hazard. Mater.* **146**, 270–277 (2007).

28. 28.

Bae, W., Chen, W., Mulchandani, A. & Mehra, R. K. Enhanced bioaccumulation of heavy metals by bacterial cells displaying synthetic phytochelatins. *Biotechnol. Bioeng.* **70**, 518–524 (2000).

29. 29.

Becher, I. et al. Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat. *Nat. Chem. Biol.* **12**, 908–910 (2016).

30. 30.

Franken, H. et al. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **10**, 1567–1593 (2015).

31. 31.

Brochado, A. R. et al. Species-specific activity of antibacterial drug combinations. *Nature* **559**, 259–263 (2018).

32. 32.

Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* **24**, 40–49 (2014).

33. 33.

Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).

34. 34.

Zhang, F. et al. *Caenorhabditis elegans* as a model for microbiome research. *Front. Microbiol.* **8**, 485 (2017).

35. 35.

Vetizou, M. et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350**, 1079–107 (2015).

36. 36.

Wu, H. et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23**, 850–858 (2017).

37. 37.

Macedo, D. et al. Antidepressants, antimicrobials or both? Gut microbiota dysbiosis in depression and possible implications of the antimicrobial effects of antidepressant drugs for antidepressant effectiveness. *J. Affect. Disord.* **208**, 22–32 (2017).

38. 38.

Sharon, G., Sampson, T. R., Geschwind, D. H. & Mazmanian, S. K. The central nervous system and the gut microbiome. *Cell* **167**, 915–932 (2016).

39. 39.

Dent, R. et al. Changes in body weight and psychotropic drugs: a systematic synthesis of the literature. *PLoS ONE* **7**, e36889 (2012).

40. 40.

Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

41. 41.

Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

42. 42.

Elias, J. E. & Gygi, S. P. Target–decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).

43. 43.

Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

44. 44.

Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

45. 45.

Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).

46. 46.

Porollo, A. EC2KEGG: a command line tool for comparison of metabolic pathways. *Source Code Biol. Med.* **9**, 19 (2014).

47. 47.

Mateus, A. et al. Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol. Syst. Biol.* **14**, e8242 (2018).

48. 48.

Hughes, C. S. et al. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).

49. 49.

Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).

50. 50.

Ortmayr, K., Charwat, V., Kasper, C., Hann, S. & Koellensperger, G. Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems. *Analyst*, **142**, 80–90 (2017).

51. 51.

He, L., Diedrich, J., Chu, Y. Y. & Yates, J. R. 3rd Extracting accurate precursor information for tandem mass spectra by RawConverter. *Anal. Chem.* **87**, 11361–11367 (2015).

52. 52.

Mahieu, N. G., Genenbacher, J. L. & Patti, G. J. A roadmap for the XCMS family of software solutions in metabolomics. *Curr. Opin. Chem. Biol.* **30**, 87–93 (2016).

53. 53.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).

54. 54.

Vinaixa, M. et al. A guideline to univariate statistical analysis for lc/ms-based untargeted metabolomics-derived data. *Metabolites* **2**, 775–795 (2012).

55. 55.

Smith, C. A. et al. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).

56. 56.

Tanabe, M. & Kanehisa, M. Using the KEGG database resource. *Curr. Protoc. Bioinformatics* **38**, 1.12.1–1.12.43 (2012).

57. 57.

Führer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry. *Anal. Chem.* **83**, 7074–7080 (2011).

58. 58.

Ponomarova, O. et al. Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow. *Cell Syst.* **5**, 345–357 (2017).

59. 59.

Wishart, D. S. et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).

60. 60.

Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).

61. 61.

Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).

62. 62.

Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

63. 63.

Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).

64. 64.

Kanehisa, M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).

## Acknowledgements

This project was supported by the European Union’s Horizon 2020 research and innovation programme under the grant agreement number 686070, and by the UK Medical Research Council (project number MC\_UU\_00025/11). A.M., L.M., M.T. and V.P. were supported by the EMBL interdisciplinary postdoctoral program. We thank EMBL Genomics, Metabolomics and Proteomics core facilities for their support in respective analyses.

## Author information

### Author notes

1. Martina Klünemann

Present address: Evonik Operations GmbH, Essen, Germany

2. Sergej Andrejev & Melanie Tramontano

Present address: German Cancer Research Center, Heidelberg, Germany

3. Manuel Banzhaf

Present address: School of Biosciences, University of Birmingham, Birmingham, UK

4. Marie-Therese Mackmull

Present address: ETH Zürich, Zürich, Switzerland

5. Leo Nesme

Present address: Molecular Health GmbH, Heidelberg, Germany

6. Ana Rita Brochado

Present address: University of Würzburg, Würzburg, Germany

7. Lisa Maier

Present address: University of Tübingen, Tübingen, Germany

8. Thomas Bock

Present address: Biozentrum, University of Basel, Basel, Switzerland

9. Carsten Schultz

Present address: Chemical Physiology and Biochemistry Department, Oregon Health & Science University, Portland, OR, USA

10. Martin Beck

Present address: Max Planck Institute of Biophysics, Frankfurt am Main, Germany

11. Janosch Hennig

Present address: Biophysical Chemistry Department, University of Bayreuth, Bayreuth, Germany

12. Filipe Cabreiro

Present address: CECAD, University of Cologne, Köln, Germany

13. These authors contributed equally: Martina Klünemann, Sergej Andrejev, Sonja Blasche, Andre Mateus

## Affiliations

### 1. European Molecular Biology Laboratory, Heidelberg, Germany

Martina Klünemann, Sergej Andrejev, Sonja Blasche, Andre Mateus, Prasad Phapale, Saravanan Devendran, Bernd Simon, Dimitrios Konstantinidis, Katharina Zirngibl, Eleonora Mastrorilli, Manuel Banzhaf, Marie-Therese Mackmull, Felix Hövelmann, Leo Nesme, Ana Rita Brochado, Lisa Maier, Thomas Bock, Vinita Periwal, Manjeet Kumar, Yongkyu Kim, Melanie Tramontano, Carsten Schultz, Martin Beck, Janosch Hennig, Michael Zimmermann, Mikhail M. Savitski, Peer Bork, Athanasios Typas & Kiran R. Patil

### 2. Medical Research Council Toxicology Unit, Cambridge, UK

Sonja Blasche, Eleni Kafkia, Katharina Zirngibl, Vinita Periwal & Kiran R. Patil

### 3. Cellzome, GlaxoSmithKline R&D, Heidelberg, Germany

Johanna Vappiani & Daniel C. Sévin

### 4. Institute of Structural and Molecular Biology, University College London, London, UK

Timothy A. Scott & Filipe Cabreiro

### 5. Institute of Clinical Sciences, Imperial College London, London, UK

Filipe Cabreiro

### 6. Max Delbrück Centre for Molecular Medicine, Berlin, Germany

Peer Bork

### 7. Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea

Peer Bork

8. Department of Bioinformatics, Biocenter, University of Würzburg,  
Würzburg, Germany

Peer Bork

## Contributions

M. Klünemann, P.B., A.T. and K.R.P. conceived the study. M. Klünemann, S.A., A.T. and K.R.P. planned the overall experiments. S.A. performed the overall data analysis. K.Z. and V.P. performed the drug clustering. M. Klünemann carried out the interaction screen, large-volume validation and UPLC data analysis. A.R.B., L.M., M.T. and M. Banzhaf assisted with the screen set-up. F.H. and C.S. designed and synthesized the clickable drug. M. Klünemann, M.-T.M. and T.B. performed the click chemistry proteomics experiments. M. Beck designed and supervised the click chemistry proteomics analysis. A.M. and M.M.S. planned the TPP experiments. S.B. and A.M. performed the TPP experiments. A.M., M.M.S. and S.A. analysed the TPP data. J.V. and D.C.S. performed the FIA–MS experiments and data analysis. S.B. performed the bacterial culturing experiments and P.P. performed the LC–MS analysis for drug measurements. M. Klünemann and P.P. performed the secreted metabolite LC–MS analysis in buffer. M. Klünemann and S.B. prepared the samples for, and B.S., L.N. and J.H. performed, the NMR analysis. M. Klünemann and D.K. performed growth assays and sample preparation for cross-fed metabolite analysis. S.B. performed growth assays and sample preparation for secreted metabolite analysis in GMM. S.D., E.M., E.K. and M.Z. performed the HILIC–MS/MS analysis. K.Z. analysed the cross-feeding metabolomics data. M. Kumar performed the motif analysis. M. Klünemann performed the community assembly experiments and Y.K. analysed the 16S data. T.A.S. and F.C. performed the *C. elegans* experiments and data analysis, D.K. measured drug concentrations. M. Klünemann and K.R.P. wrote the paper.

## Corresponding authors

Correspondence to Peer Bork or Athanasios Typas or Kiran R. Patil.

# Ethics declarations

## Competing interests

M. Klünemann, S.A., L.M., M.T., Y.K., P.B., A.T. and K.R.P. are inventors in a patent application based on the findings reported in this study (US patent application number 16966322). S.B., A.M., P.P., S.D., J.V., B.S., T.A.S., E.K., D.K., K.Z., E.M., M. Banzhaf, M.-T.M., F.H., L.N., A.R.B., T.B., V.P., M. Kumar, C.S., M. Beck, J.H., M.Z., D.C.S., F.C. and M.M.S. declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Kim Lewis, Michael Shapira and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Bacteria and drug selection.

**a**, Distribution of the selected 25 bacterial strains by their phylogenetic class, and their cumulative metabolic diversity measured as the coverage of annotated enzymes as per the KEGG database<sup>64</sup>. **b**, We started with approximately 1,000 annotated drugs from the SIDER side effect database (Kuhn et al. 2016), which were filtered for their gut related side effects. Drug selection was enriched from another database (Saad et al. 2012) for known or suspected interactions with the gut microbiome, before filtered for oral administration and manually curated for overall interest. Final selection was filtered for availability from vendors and establishment of UPLC methods. **c**, The drugs used in this study span a broad range of structural diversity. Shown is the spread of the selected drugs in a principle

coordinate analysis, covering >2,000 drugs from the DrugBank database. Maximum common sub-structure was used to calculate the distances between drug pairs. **d**, Selected drugs cover several therapeutic classes / indication areas. **e**, Chemical structures of the 15 drugs used in this study.

**Extended Data Fig. 2 Correlation between the screen and validation in higher-volume cultures.**

For screen,  $n \geq 4$  independent replicates (median number of replicates = 17). For validation,  $n = 3$ . Error bars = S.E.M. For screening, multiple independent batches were performed as indicated in Supplementary Table 3. Shown R (correlation coefficient) and p-value based on Pearson correlation test.

**Extended Data Fig. 3 NMR measurements showing duloxetine depletion by bacterial cells.**

**a**, *B. uniformis*, **b**, *E. coli ED1A*, **c**, *E. coli IAI1*, and **d**, *C. saccharolyticum*. **e**, NMR spectrum from *C. saccharolyticum* cell pellet extract showing that the recovered drug is unmodified duloxetine. Resonances appearing to be out of phase and strong baseline distortions are due to the presence of large solvent signals outside the displayed chemical shift range.

**Extended Data Fig. 4 NMR measurements showing unmodified duloxetine recovered from bacterial pellet.**

Bacterial cells were incubated with the drug for 4 h in PBS buffer prior to recovery. **a**, Illustration of the experimental procedure marking the sample collection points. **b**, NMR spectra of recovered duloxetine from different fractions of *E. coli* IAI1 and *C. saccharolyticum* preincubated in PBS. The reference spectrum was scaled to the amount present in the sample to assess the relative amount of free duloxetine present in respective samples. Resonances appearing to be out of phase and strong baseline distortions are due to the presence of large solvent signals outside the displayed chemical shift range.

## Extended Data Fig. 5 Duloxetine bioaccumulation by *E. coli* IAI1 in GMM and recovery from pellet.

**a**, Procedure and collected samples: S0-S4. **b**, Recovered duloxetine from different samples (S0-S4) collected as described in **a**. Different starting duloxetine concentrations, between 0-70  $\mu$ M, were used. S0 = medium without bacteria (drug only control), S1 = total culture (medium plus bacteria), S2 = supernatant, S3 = wash (pellet was washed with PBS, no drug was found therein supporting intracellular accumulation), S4 = washed pellet. n = 3, error bars = SD, central squares mark the mean. **c**, MS/MS spectra of duloxetine standard (bottom) and duloxetine detected in a S1 sample (top).

## Extended Data Fig. 6 Molecular effects of duloxetine bioaccumulation.

**a**, Alkynated duloxetine made for the biotin-pull down assay. **b**, Fold change of proteins detected in the duloxetine pull down assay in *C. saccharolyticum* lysate using alkynated duloxetine. Four replicates were used in both test and control sets. Significantly enriched (hypergeometric test, FDR corrected  $p < 0.1$ ,  $\log_2(Fc) > 2$ ) proteins are shown in red. **c**, **d**, Bioaccumulating *E. coli* strain features larger change in protein abundance in response to drug treatment. Shown are the number of proteins with altered abundance in *E. coli* ED1A (**c**, non-bioaccumulating), and *E. coli* IAI1 (**d**, bioaccumulating) strains in response to duloxetine exposure at different concentrations. **e–h**, Comparison of MS/MS spectra of four nucleotide-pathway metabolites from the supernatant of duloxetine-treated *C. saccharolyticum* with MS/MS spectra of analytical standards. (CE = 10 eV; further details in Methods) Related to Fig. 2d and Supplementary Table 11.

## Extended Data Fig. 7 Duloxetine induces a shift in metabolite secretion.

**a**, Effect of duloxetine treatment on the exo-metabolome of six gut bacterial strains. Shown is the distribution of individual samples over the first two

principle components. Principle Component Analysis (PCA) was performed on untargeted FIA–MS data (Methods). The numbers in parentheses of PC1 and PC2 mark the corresponding explained variance for the first and the second principle component, respectively. The dotted block arrow marks the duloxetine induced shift in exo-metabolome of *C. saccharolyticum*. **b**, Duloxetine concentration dependent changes in the *C. saccharolyticum* exo-metabolome. The ion mapping to the deprotonated duloxetine was removed from the PCA analysis shown in **a** and **b**. **c**, The signal for the closest matching ion for deprotonated duloxetine [M-H]<sup>-</sup> from the exometabolomics data (*m/z* 296.110079) plotted against initial duloxetine concentration. Data from all six species are pooled together (*n* = 24 for each initial duloxetine concentration). Overlaid box plots show the interquartile range (IQR), the median value and whiskers extending to include all the values less than  $1.5 \times$  IQR away from the 1st or 3rd quartile, respectively. **d**, Duloxetine signal in the FIA–MS data stratified by species. The signal for the closest matching ion for deprotonated duloxetine [M-H]<sup>-</sup> from the exometabolomics data (FIA–MS) (*m/z* 296.110079) plotted against initial duloxetine concentration. Thick transparent line traces medians of replicates (*n* = 4) at each initial concentration. The dotted lines show linear regression fit.

### Extended Data Fig. 8 Duloxetine-induced exo-metabolome changes.

**a**, Change in *C. saccharolyticum* exo-metabolome (HILIC-MS data) in response to non-bioaccumulated roflumilast. **b**, Same as in Fig. 2g, but based on 69 metabolites, whose chemical identity was putatively assigned, and confirmed for 2 metabolites using chemical standards (Supplementary Fig. 3, Supplementary Table 17), using HILIC-MS/MS analysis.

### Extended Data Fig. 9 Duloxetine bioaccumulation, community assembly and host response.

**a**, *E. rectale* relative abundance in transfer assays based on 16-S amplicon reads. **b**, *E. rectale* relative abundance as in **a** but normalized with respect to equal abundance of each of the five species in the inoculum mixture.

Mean values from biological triplicates are shown. **c**, Duloxetine depletion in community assembly assay. Dashed line indicates mean of control. n = 6 (3 biological replicates, 2 measurements per sample); overlaid box plots show the interquartile range (IQR), the median value and whiskers extending to include all the values less than  $1.5 \times$  IQR away from the 1st or 3rd quartile, respectively. **d**, Metabolic cross-feeding between *S. salivarius* and *E. rectale*. Shown are the results of untargeted metabolomics analysis (FIA–MS) of supernatants collected during the growth of *S. salivarius* in GMM with duloxetine and the subsequent growth of *E. rectale* in the cell-free conditioned medium. Shown are the profiles of the ions that increased during *S. salivarius* growth and decreased during *E. rectale* growth, implying cross-feeding. Ions showing similar pattern in the drug-free solvent (DMSO) control were filtered out. Mean intensities from three biological replicates are shown. **e**, Dose dependent effects of duloxetine on muscular function in wild type *C. elegans* animals. Larval stage four (L4) worms were incubated in LB medium in the presence of duloxetine at the indicated concentrations. Each bar represents the mean of six independent experiments, each performed with two technical replicates,  $\pm$  SD. P values mark difference to the no-drug control, estimated using one-way ANOVA followed by correction for multiple pair-wise comparisons (Tukey's test). **f**, Duloxetine concentration in the *C. elegans* behaviour assay (n = 6; 3 biological replicates, 2 measurements per sample). Overlaid box plots show the interquartile range (IQR), the median value and whiskers extending to include all the values less than  $1.5 \times$  IQR away from the 1st or 3rd quartile, respectively

[Source data](#).

## Supplementary information

### [Supplementary Information](#)

This file contains Supplementary Figures 1–3 and the legends from Supplementary Tables 1–18.

### [Reporting Summary](#)

## **Supplementary Tables 1–7**

See main Supplementary Information PDF for table legends.

## **Supplementary Table 8**

See main Supplementary Information PDF for table legend.

## **Supplementary Tables 9–18**

See main Supplementary Information PDF for table legends.

## **Source data**

### **Source Data Fig. 2**

### **Source Data Fig. 3**

### **Source Data Extended Data Fig. 9**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Klünemann, M., Andrejev, S., Blasche, S. *et al.* Bioaccumulation of therapeutic drugs by human gut bacteria. *Nature* **597**, 533–538 (2021). <https://doi.org/10.1038/s41586-021-03891-8>

- Received: 25 February 2019
- Accepted: 10 August 2021

- Published: 08 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03891-8>

## 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/s41586-021-03891-8>

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

- Article
- [Published: 15 September 2021](#)

# Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks

- [Alpha Kabinet Keita](#) [ORCID: orcid.org/0000-0003-4377-341X<sup>1,2</sup>](#) [na1](#),
- [Fara R. Koundouno<sup>3,4</sup>](#) [na1](#),
- [Martin Faye<sup>5</sup>](#) [na1](#),
- [Ariane Dux<sup>6</sup>](#) [na1](#),
- [Julia Hinzmann<sup>4,7,8</sup>](#) [na1](#),
- [Haby Diallo<sup>1</sup>](#),
- [Ahidjo Ayouba<sup>2</sup>](#),
- [Frederic Le Marcis](#) [ORCID: orcid.org/0000-0001-8302-0864<sup>1,2,9</sup>](#),
- [Barré Soropogui<sup>3</sup>](#),
- [Kékoura Ifono<sup>3,4</sup>](#),
- [Moussa M. Diagne<sup>5</sup>](#),
- [Mamadou S. Sow<sup>1,10</sup>](#),
- [Joseph A. Bore<sup>3,11</sup>](#),
- [Sebastien Calvignac-Spencer](#) [ORCID: orcid.org/0000-0003-4834-0509<sup>6</sup>](#),
- [Nicole Vidal<sup>2</sup>](#),
- [Jacob Camara<sup>3</sup>](#),
- [Mamadou B. Keita<sup>12</sup>](#),
- [Annick Renevey<sup>4,7</sup>](#),
- [Amadou Diallo<sup>5</sup>](#),
- [Abdoul K. Soumah<sup>1</sup>](#),
- [Saa L. Millimono<sup>3,4</sup>](#),
- [Almudena Mari-Saez<sup>6</sup>](#),
- [Mamadou Diop<sup>5</sup>](#),
- [Ahmadou Doré<sup>3</sup>](#),
- [Fodé Y. Soumah<sup>10</sup>](#),
- [Kaka Kourouma<sup>12</sup>](#),
- [Nathalie J. Vielle<sup>4,13</sup>](#),

- [Cheikh Loucoubar](#)<sup>5</sup>,
- [Ibrahima Camara](#)<sup>1</sup>,
- [Karifa Kourouma](#)<sup>3,4</sup>,
- [Giuditta Annibaldi](#)<sup>4,13</sup>,
- [Assaïtou Bah](#)<sup>3</sup>,
- [Anke Thielebein](#)<sup>4,7</sup>,
- [Meike Pahlmann](#)<sup>4,7</sup>,
- [Steven T. Pullan](#)<sup>8,11</sup>,
- [Miles W. Carroll](#)<sup>8,11</sup>,
- [Joshua Quick](#)<sup>14</sup>,
- [Pierre Formenty](#) ORCID: [orcid.org/0000-0002-9482-5411](http://orcid.org/0000-0002-9482-5411)<sup>15</sup>,
- [Anais Legand](#)<sup>15</sup>,
- [Karla Pietro](#)<sup>16</sup>,
- [Michael R. Wiley](#)<sup>16,17</sup>,
- [Noel Tordo](#)<sup>18</sup>,
- [Christophe Peyrefitte](#)<sup>5</sup>,
- [John T. McCrone](#) ORCID: [orcid.org/0000-0002-9846-8917](http://orcid.org/0000-0002-9846-8917)<sup>19</sup>,
- [Andrew Rambaut](#) ORCID: [orcid.org/0000-0003-4337-3707](http://orcid.org/0000-0003-4337-3707)<sup>19</sup>,
- [Youssouf Sidibé](#)<sup>20</sup>,
- [Mamadou D. Barry](#)<sup>20</sup>,
- [Madeleine Kourouma](#)<sup>20</sup>,
- [Cé D. Saouromou](#)<sup>20</sup>,
- [Mamadou Condé](#)<sup>20</sup>,
- [Moussa Baldé](#)<sup>10</sup>,
- [Moriba Povogui](#)<sup>1</sup>,
- [Sakoba Keita](#)<sup>21</sup>,
- [Mandiou Diakite](#)<sup>22,23</sup>,
- [Mamadou S. Bah](#)<sup>22</sup>,
- [Amadou Sidibe](#)<sup>9</sup>,
- [Dembo Diakite](#)<sup>10</sup>,
- [Fodé B. Sako](#)<sup>10</sup>,
- [Fodé A. Traore](#)<sup>10</sup>,
- [Georges A. Ki-Zerbo](#)<sup>13</sup>,
- [Philippe Lemey](#) ORCID: [orcid.org/0000-0003-2826-5353](http://orcid.org/0000-0003-2826-5353)<sup>24</sup>,
- [Stephan Günther](#) ORCID: [orcid.org/0000-0002-6562-0230](http://orcid.org/0000-0002-6562-0230)<sup>4,7,13</sup>,
- [Liana E. Kafetzopoulou](#)<sup>4,7,24</sup>,
- [Amadou A. Sall](#)<sup>5</sup>,
- [Eric Delaporte](#)<sup>2,25</sup>,
- [Sophie Duraffour](#)<sup>4,7,13 na2</sup>,

- [Ousmane Faye<sup>5 na2</sup>](#),
- [Fabian H. Leendertz](#) [ORCID: orcid.org/0000-0002-2169-7375<sup>6 na2</sup>](#),
- [Martine Peeters<sup>2 na2</sup>](#),
- [Abdoulaye Toure<sup>1,12 na2</sup>](#) &
- [N'. Faly Magassouba<sup>3 na2</sup>](#)

[Nature](#) volume **597**, pages 539–543 (2021)

- 12k Accesses
- 732 Altmetric
- [Metrics details](#)

## Subjects

- [Ebola virus](#)
- [Epidemiology](#)

## Abstract

Seven years after the declaration of the first epidemic of Ebola virus disease in Guinea, the country faced a new outbreak—between 14 February and 19 June 2021—near the epicentre of the previous epidemic<sup>1,2</sup>. Here we use next-generation sequencing to generate complete or near-complete genomes of *Zaire ebolavirus* from samples obtained from 12 different patients. These genomes form a well-supported phylogenetic cluster with genomes from the previous outbreak, which indicates that the new outbreak was not the result of a new spillover event from an animal reservoir. The 2021 lineage shows considerably lower divergence than would be expected during sustained human-to-human transmission, which suggests a persistent infection with reduced replication or a period of latency. The resurgence of *Zaire ebolavirus* from humans five years after the end of the previous outbreak of Ebola virus disease reinforces the need for long-term medical and social care for patients who survive the disease, to reduce the risk of re-emergence and to prevent further stigmatization.

[Download PDF](#)

## Main

At least 30 outbreaks of Ebola virus disease (EVD) have been identified since the late 1970s, the most severe of which affected Guinea, Sierra Leone and Liberia from

December 2013 to June 2016<sup>1,2</sup>. Guinea experienced a new outbreak of EVD in 2021, which started in Gouéké—a town about 200 km away from the epicentre of the 2013–2016 outbreak. The probable index case was a 51-year-old nurse, an assistant of the hospital midwife in Gouéké. On 21 January 2021, she was admitted to hospital in Gouéké suffering from headache, asthenia, nausea, anorexia, vertigo and abdominal pain. She was diagnosed with malaria and salmonellosis and was released two days later. Feeling ill again once at home, she attended a private clinic in Nzérékoré (40 km away) and visited a traditional healer, but died three days later. In the week after her death, her husband—as well as other family members who attended her funeral—fell ill, and four of them died. They were reported as the first suspect cases by the national epidemic alert system on 11 February. On 12 February, blood was taken from two suspect cases admitted to hospital in Nzérékoré. On 13 February, both of these patients were confirmed to have EVD by the laboratory in Guéckédou, which used a commercial real-time polymerase chain reaction with reverse transcription (qRT–PCR) assay (RealStar Filovirus Screen Kit, Altona Diagnostics). On 13 February, the husband of the index case—who travelled more than 700 km from Gouéké to Conakry, the capital city of Guinea, for treatment—was admitted to the Centre de Traitement Epidémiologique (CTEpi) in Nongo, Ratoma Commune. He presented with fever, nausea, asthenia, abdominal pain and lumbar pain and was strongly suspected to have EVD. A blood sample was analysed on the same day and was found to be positive for Ebola Zaire (*Zaire ebolavirus*; EBOV) according to the GeneXpert molecular diagnostic platform (Xpert Ebola test, Cepheid) and by an in-house qRT–PCR assay. Laboratory confirmation of EVD in the three suspect cases led to the official declaration of the epidemic on 14 February. By 5 March, 14 confirmed cases and 4 probable cases of EVD had been identified, leading to 9 deaths—including 5 confirmed cases as reported by the Agence Nationale de la Sécurité Sanitaire (ANSS) of Guinea. After a period of 25 days without new cases, two new cases were reported around Nzérékoré on 1 and 3 April, and on 19 June 2021 the outbreak was declared to be over. In total, 16 confirmed cases were reported, among which 12 people died.

Genomic characterization of the virus that caused the 2021 epidemic of EVD in Guinea was of immediate importance to public health. First, because diagnostic tools, therapeutics and vaccines with proven effectiveness in recent EVD outbreaks—such as in Guinea (2013–2016) and in the Equateur and North-Kivu/Ituri provinces of the Democratic Republic of the Congo (DRC) (2018–2020)—have primarily been developed for EBOV<sup>3,4,5</sup>. Second, to identify whether the outbreak resulted from a new zoonotic transmission event or from the resurgence of a viral strain that had circulated in a previous EBOV outbreak: it is known that EBOV can persist in the bodily fluids of patients who have survived EVD and can be at the origin of new transmission chains<sup>6,7,8</sup>. Although the Xpert Ebola test was developed to detect only EBOV strains and the in-house qRT–PCR assay uses a probe that is specifically designed to detect EBOV<sup>9</sup>, additional confirmation by sequence analysis was sought

by targeting a short fragment in the viral protein 35 region of the sample from the patient who was hospitalized in Conakry. The phylogenetic tree (Supplementary Fig. 1) underscores that this highly conserved region can discriminate between Ebola virus species, and analysis confirmed that the virus that caused the new outbreak was of the species *Zaire ebolavirus*. This confirmed that available vaccines and the vast majority of molecular-diagnostic tools and therapeutics could be immediately applied.

To gain further insight into the genomic make-up of the viruses causing this outbreak, 11 complete or near-complete (greater than 95% recovery) and 8 partial (greater than 65% recovery) genomic sequences from 12 of the 14 confirmed cases were obtained by 3 different laboratories using different next-generation sequencing technologies (Table 1). To facilitate the public-health response and the evaluation of existing medical countermeasures, sequencing results were made publicly available on 12 March through joint posting (<https://virological.org/c/ebolavirus/guinea-2021/44>). Blood and swab samples from 14 patients with confirmed EVD, sampled from 12 February to 4 March, were processed by the following methods: hybridization capture technology and sequencing on Illumina iSeq100, an amplicon-based protocol with EBOV-specific primer pools and sequencing on MinION (Oxford Nanopore Technologies (ONT)), and a hybrid-capture-based approach using a probe panel that included EBOV-specific targets followed by TruSeq exome enrichment, as previously described<sup>5</sup>. The data generated between the three groups were pooled and the sequence that had the highest quality was chosen for each patient. This enabled us to reconstruct 12 high-quality EBOV genomes that covered 82.9–99.9% of the reference genome (KR534588) (Table 1). The consensus EBOV sequences with the highest genome recovery (greater than 82.9%) from 12 different patients were used in further analyses.

**Table 1 Patient and sample characteristics and sequencing results obtained by the laboratories involved in the study**

Maximum likelihood phylogenetic reconstruction places the 12 genomes from the 2021 outbreak of EVD in Guinea as a single cluster among the EBOV viruses that were responsible for the 2013–2016 outbreak in West Africa (Figs. 1, 2). The genomes from the 2021 outbreak share 10 substitutions (compared with KJ660346) that were accumulated during the 2013–2016 outbreak, including the A82V marker mutation for human adaptation in the glycoprotein that arose when the virus spread to Sierra Leone<sup>11,12</sup>. These patterns provide strong evidence of a direct link to human cases from the 2013–2016 outbreak rather than a new spillover from an animal reservoir. The 2021 lineage is nested within a clade that predominantly consists of genomes sampled from Guinea in 2014 (Fig. 2). The branch by which the 2021 cluster diverges from the previous outbreak exhibits only 12 substitutions, which is far fewer than would be expected from the evolution of EBOV during 6 years of sustained human-to-human transmission (Fig. 3). Using a local molecular-clock analysis, we estimate a

6.4-fold (95% highest posterior density (HPD) interval: 3.3-fold, 10.1-fold) lower rate along this branch. For comparison, we also estimate a 5.5-fold (1.6-fold, 10.8-fold) lower rate along the branch leading to the 2016 cases, which were linked to a patient who survived the disease and in whom the virus persisted for more than 500 days<sup>7,13</sup>. Rather than a constant long-term low evolutionary rate, some degree of latency or dormancy during persistent infection seems to be a more likely explanation for the low divergence of the genomes from the 2021 epidemic. We tested whether the 12 genomes from the 2021 epidemic, which were sampled over a time period of less than one month, contained sufficient temporal signal to estimate the time to most recent common ancestor (tMRCA) (Supplementary Fig. 2); however, we did not identify statistical support for sufficient divergence accumulation over this short timescale. We therefore calibrated our analysis using an evolutionary rate that reflects EBOV evolution under sustained human-to-human transmission (as estimated by the local molecular-clock analysis). This resulted in a tMRCA estimate of 22 January 2021 (95% HPD interval: 29 December 2020, 10 February 2021).

**Fig. 1: Maximum likelihood phylogenetic reconstruction for 55 representative genomes from previous outbreaks of *Zaire ebolavirus* and 12 genomes from the 2021 outbreak in Guinea.**

 figure1

Most clades for single or multiple closely related outbreaks are collapsed and internal node support is proportional to the size of the internal node circles. The clades or tip circles are labelled with the locations and years of the outbreaks, and coloured according to the (first) year of detection.

**Fig. 2: Maximum likelihood phylogenetic reconstruction for 1,065 genomes sampled during the 2013–2016 West African outbreak and 12 genomes from the 2021 outbreak in Guinea.**

 figure2

A colour gradient (from purple to green for increasing divergence) is used to colour the tip circles. The 2021 genomes are shown with a larger circle in yellow.

**Fig. 3: Temporal divergence plot showing genetic divergence from the root over time.**

---

 **figure3**

---

This plot relates to the tree shown in Fig. 2. The regression is exclusively fitted to genomes sampled between 2014 and 2015. The same colours are used for the data points as in Fig. 2. The dashed yellow lines highlight how the 2021 data points deviate from the relationship between sampling time and sequence divergence. According to this relationship, about 95 substitutions (95% prediction interval: 88–101) are expected on the branch ancestral to the 2021 cluster, whereas only 12 are inferred on this branch.

These results open up a new perspective on the relatively rare observation of EBOV re-emergence. It is assumed that all known filovirus outbreaks in humans are the result of independent zoonotic transmission events from bat reservoir species or from intermediate or amplifying hosts such as apes and duikers<sup>6</sup>. Here we clearly show that, even almost five years after the declaration of the end of an epidemic, new outbreaks could also be the result of transmission from humans who were infected during a previous epidemic. The viruses from the 2021 outbreak fall within the lineage of EBOV viruses obtained from humans during the 2014–2016 outbreak; as such, it is very unlikely that this new outbreak has an animal origin or is the result of a new cross-species transmission with the same lineage that remained latent in this natural host, which in that scenario would be at the basis of the West African cluster. The limited genomic divergence between 2014–2015 and 2021 is compatible with a slow long-term evolutionary rate. However, a relatively long phase of latency might be more likely than continuous slow replication. Independent of the mechanistic explanation, the virus most probably persisted at a low level in a human who had survived previous infection. Plausible scenarios of EBOV transmission to the index case include: sexual transmission by exposure to EBOV in semen from a male survivor; contact with body fluids from a survivor who had a relapse of symptomatic EVD (for example during healthcare—the index case was a healthcare worker); or

relapse of EVD in the index case—although she was not known to have been infected previously, she could have had an asymptomatic or pauci-symptomatic EBOV infection during the previous outbreak. A detailed investigation of the family of the index case by anthropologists revealed that she was not known to have had EVD previously, nor were her husband or close relatives. However, among more distantly related family, 25 individuals had EVD during the previous outbreak. Only five survived, although the index case apparently had no recent contacts with this part of the family. Consultation of the hospital registers in Gouécké showed that all patients seen by the index case in January 2021 were in good health and were still in good health in March 2021. However, the index case also performed informal consultations outside the hospital environment, which could not be verified. An alternative scenario is that the nurse was not the actual index case, but was part of a small, unrecognized chain of human-to-human transmission in this area of Guinea. However, the diversity of the currently available genomes is limited, and molecular-clock analysis suggests a recent tMRCA, with a mean estimate close to the time that the nurse was first hospitalized and a 95% HPD boundary around the beginning of the year. This provides some reassurance that the outbreak was detected early.

The 2013–2016 outbreak in West Africa was the largest and most complex recent outbreak of EBOV, and involved more than 28,000 cases, 11,000 deaths and an estimated 17,000 survivors, mostly in Guinea, Liberia and Sierra Leone<sup>2</sup>. This large outbreak provided new information about the disease itself as well as about the medical, social and psychological implications for patients who survived the disease<sup>14,15,16</sup>. It was also possible to estimate, to some extent, the proportions of asymptomatic or pauci-symptomatic infections and to identify their role in specific unusual transmission chains<sup>17,18,19</sup>. Although the main route of human-to-human transmission of EBOV is direct contact with infected bodily fluids from symptomatic or deceased patients, some transmission chains in this outbreak were associated with viral persistence in semen<sup>3</sup>. Several studies demonstrated viral persistence in more than 50% of male survivors at 6 months after discharge from Ebola treatment units (ETU), and the maximum duration of persistence in semen has been reported to be up to 500–700 days after ETU discharge in a small number of male EVD survivors<sup>9,20,21,22</sup>. Transmission through other bodily fluids (such as breast milk and cervicovaginal fluids) is also suspected<sup>8,23,24,25</sup>. Furthermore, some immunological studies among survivors suggest a continuous or intermittent EBOV antigenic stimulation due to persistence of an EBOV reservoir in some survivors<sup>26,27</sup>, although this was not confirmed in another study<sup>28</sup>. Cases of relapse of EVD have also been sporadically reported and could be the origin of large transmission chains, as recently reported in the North-Kivu outbreak in DRC<sup>29</sup>. For example, the presence of EBOV RNA, 500 days after ETU discharge, in the breast milk of a woman who was not pregnant when she developed EVD has recently been reported. She attended the hospital owing to complications at 8 months of pregnancy, and a breast milk sample

that was taken 1 month after delivery tested positive for EBOV RNA<sup>9</sup>. These examples illustrate that healthcare workers can be exposed to EBOV when taking care of patients who survived EVD but have an unrecognized relapse of their infection. The 2021 outbreak now highlights that viral persistence and reactivation is not limited to a two-year period, but can also occur on much longer timescales with late reactivation.

Active genomic surveillance has already shown the resurgence of previous strains in other outbreaks of the disease. For example, two EBOV variants circulated simultaneously within the same region during the recent 2020 outbreak in Equateur province, DRC<sup>30</sup>. Moreover, strains from the two consecutive outbreaks in Luebo, DRC, in 2007 and 2008, are also so closely related that it now seems difficult to exclude that the epidemic observed in 2008 was due to a resurgence event from patient who survived EVD in the 2007 outbreak<sup>31,32</sup>. However, the limited genomic sampling does not allow for a formal test of this hypothesis.

Although the majority of EVD outbreaks remained limited both in the number of cases and in geographic spread, the two largest outbreaks in West Africa (December 2013–June 2016) and in eastern DRC (August 2018–June 2020) infected thousands of individuals over wide geographic areas, leading to large numbers of EVD survivors. This means that the risk of resurgence is higher than ever before. Continued surveillance of EVD survivors is therefore warranted to monitor the reactivation and relapse of EVD infection and the potential presence of the virus in bodily fluids. This work and associated communications must be conducted with the utmost care for the wellbeing of EVD survivors. During the 2013–2016 outbreak in Guinea, patients who survived EVD had a mixed experience after discharge from ETUs. On the one hand, they were considered as heroes by non-governmental organizations and became living testimonies of a possible recovery<sup>33,34</sup>. On the other hand, they experienced different forms of stigmatization, such as rejection by family and friends, refusal of involvement in collective work, loss of jobs and housing, and sometimes self-isolation from social life and workplaces<sup>35</sup>. The human origin of the 2021 EVD outbreak, and the associated shift in our perception of EBOV emergence, call for careful attention to survivors of the disease. The concern that survivors will be stigmatized as a source of danger should be a matter of scrupulous attention<sup>36</sup>. This is especially true for the area of Gouécké, which is only 9 km away from Womey—a village that is emblematic of the violent reaction of the population towards the EVD response team during the 2013–2016 epidemic<sup>37</sup>.

Since the 2013–2016 EVD outbreak in Western Africa, genome sequencing has become a major component of the response to outbreaks<sup>10,38,39,40,41</sup>. The establishment of in-country sequencing and the building of capacity enabled a timely characterization of EBOV strains in the 2021 outbreak in Guinea. In addition to the importance of appropriate healthcare measures focused on survivors, the late

resurgence of the virus also highlights the urgent need for further research into potent antiviral agents that can eradicate the latent virus reservoir in patients with EVD, and into efficient vaccines that provide long-term protection. In parallel, vaccination could also be considered to boost protective antibody responses in survivors of the disease<sup>27</sup>. The vaccination of populations in areas with previous EBOV outbreaks could also be promoted to prevent secondary cases.

## Methods

### Ethics statement

Diagnostic specimens were collected as part of the emergency response from the Ministry of Health of Guinea, and therefore consent for sample collection was waived. All preparation of samples for sequencing, genomic analysis and data analysis was performed on anonymized samples identifiable only by their laboratory or epidemiological identifier.

### Confirmation of Ebola virus species by sequence analysis of the VP35 fragment at CERFIG

Viral RNA was extracted from 140 µl of whole blood collected from samples from the patient hospitalized in Conakry, using the Nuclisens kit (Biomerieux) and following the manufacturer's instructions. Amplification of a small fragment of the VP35 region was attempted in a semi-nested PCR with a modified protocol as previously described<sup>4</sup>. First-round VP35 PCR products from positive samples were barcoded and pooled using the Native Barcoding Kit EXP-NBD104 (ONT). Sequencing libraries were generated from the barcoded products using the Genomic DNA Sequencing Kit SQK-LSK109 (ONT) and were loaded onto a R9 flow cell on a MinION (ONT). Genetic data were collected for 1 h. Basecalling, adapter removal and demultiplexing of .fastq files were performed with MinKNOW, v.4.1.22. Fastq reads >Q11 were used for mapping a virus database with the Genome Detective tool (<https://www.genomedetective.com/app/typingtool/virus/>). The generated consensus sequence was used for further analysis. For phylogenetic inference, we retrieved one sequence per outbreak from the haemorrhagic fever virus (HFV) database to which we added the newly generated VP35 sequence of the new outbreak. Phylogenetic analyses were performed using maximum likelihood methods using IQ-TREE with 1,000 bootstraps for branch support<sup>42,43</sup>. The general time-reversible (GTR) model plus a discrete gamma distribution were used as nucleotide substitution models.

### Full-length genome sequencing of the new Ebola viruses

## Genome sequencing at CERFIG

Whole-genome sequencing was attempted on viral extracts for samples that were positive for EBOV glycoprotein (GP) and nucleoprotein (NP) on the GeneXpert molecular diagnostic platform (Xpert Ebola Assay) with the GP and NP of *Zaire ebolavirus*. We extracted full nucleic acid using the QIAamp Viral RNA Mini Kit (Qiagen). After DNase treatment with TURBO DNA-free Kit (Ambion) and clean-up with RNA Clean & Concentrator Kit (Zymo Research), RNA was converted to double-stranded cDNA (ds-cDNA) using the SuperScript IV First-Strand Synthesis System (Invitrogen) and NEBNEXT mRNA Second Strand Synthesis Module (New England Biolabs). The resulting ds-cDNA was enzymatically fragmented with NEBNext dsDNA Fragmentase (New England Biolabs) and converted to dual indexed libraries with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) and NEBNext Multiplex Oligos for Illumina (New England Biolabs). To enrich EBOV in the libraries, we performed two rounds of hybridization capture (16 h at 65 °C) with custom-made biotinylated RNA baits (120 nucleotides, 2-fold tiling; Arbor Biosciences) covering representative genomes for *Zaire ebolavirus* (KC242801), *Sudan ebolavirus* (KC242783), *Reston ebolavirus* (NC\_004161), *Taï Forest ebolavirus* (NC\_014372), *Bundibugyo ebolavirus* (KC545395) and *Marburg marburgvirus* (FJ750956), following the myBaits Hybridization Capture for Targeted NGS protocol (v.4.01). After the second round, capture products were quantified using the Qubit 3.0 Fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen), and pooled in equimolar amounts for sequencing on an Illumina iSeq using iSeq 100 i1 Reagents (2 × 150 cycles). Sequencing reads were filtered (adapter removal and quality filtering) with Trimmomatic<sup>44</sup> (settings: LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:40), merged with ClipAndMerge (<https://github.com/apeltzer/ClipAndMerge>), and mapped to the *Zaire ebolavirus* RefSeq genome (NC\_002549) using BWA-MEM<sup>45</sup>. Mapped reads were sorted and deduplicated with SortSam and MarkDuplicates from the Picard suite (Broad Institute, Picard; <http://broadinstitute.github.io/picard>). We generated consensus sequences using Geneious Prime 2020.2.3 (<https://www.geneious.com>), in which unambiguous bases were called when at least 90% of at least 20 unique reads were in agreement (20×, 90%). For samples with few mapped reads (0001, 0002, 0010, 0030), we also called a consensus at 2×, 90% and 5×, 90%.

## Genome sequencing at PFHG

Sequencing at PFHG was performed using a mobile MinION facility deployed by BNITM to Guinea at the beginning of March 2021. A total of 13 EBOV-positive initial diagnostic samples processed at the Laboratoire des Fièvres Hémorragiques Virales de Gueckédou, the Laboratoire Régional de l'Hôpital de Nzérékoré were used for sequencing. If RNAs from diagnostic procedures performed by the peripheral

laboratories was not sent to PFHG, samples were inactivated and RNA was extracted from 50 µl for whole blood EDTA, 70 µl of plasma from EDTA blood or from 140 µl of wet swabs using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. Tiled primers generating overlapping products combined with a highly multiplexed PCR protocol were used for amplicon generation<sup>10</sup>. At start of deployment, three different primer pools (V3 or pan\_10\_EBOV, V4 or pan\_EBOV and Zaire-PHE or EBOV-Zaire-PHE) were tested and results were combined for the optimal recovery of consensus. A new primer pool V5 (EBOV-Makona-V5) was further designed and implemented to increase consensus recovery. Primer pools V3, V4 and V5 were designed by the ARTIC network and Zaire-PHE primer pools by Public Health England (PHE). For V3, 62 primers were used, while for V4 and V5, 61 primers pairs were used, to amplify products of around 400 nt in length. For Zaire-PHE, 71 primer pairs were used to amplify products of around 350 nt in length for the approximately 20 kb viral genome. All primer pools used can be found in Supplementary Table 1. The multiplex PCR was performed as described by the most up-to-date ARTIC protocol for nCoV-2019 amplicon sequencing (nCoV-2019 sequencing protocol V3 (LoCost) V.3 (<https://artic.network/ncov-2019>), adapted to include the EBOV-specific primer sets. In brief, RNA was directly used for cDNA synthesis using the LunaScript RT SuperMix (New England Biolabs) and the cDNA generated was used as template in the multiplex PCR, which was performed in two reaction pools using Q5 Hot Start DNA Polymerase (New England Biolabs). The resulting amplicons from the two PCR pools were pooled in equal volumes and the pooled amplicons were diluted 1:10 with nuclease-free water.

Sequencing libraries were prepared, barcoded and multiplexed using the Ligation Sequencing Kit (SQK-LSK109) from ONT combined with the Native Expansion pack (EXP-NDB104, EXP-NBD114, EXP-NBD196) following the ARTIC Network's library preparation protocol (nCoV-2019 sequencing protocol v3 (LoCost) V.3 (<https://artic.network/ncov-2019>)). For the preparation of fewer than 11 samples, each sample was prepared in multiples to achieve the library concentration required for sequencing. In brief, the diluted pooled amplicons were end-repaired using the Ultra II End Prep Module (New England Biolabs) followed by barcode ligation using the Blunt/TA Ligase Master Mix and one unique barcode per sample. Equal volumes from each native barcoding reaction were pooled and subject to bead clean-up using 0.4× AMPure beads. The pooled barcoded amplicons were quantified using the Qubit Fluorometer (Thermo Fisher Scientific) and AMII adapter ligation was performed using the Quick T4 DNA Ligase (New England Biolabs) followed by an additional bead clean-up. The adaptor-ligated barcoded amplicon pool was quantified using the Qubit Fluorometer (Thermo Fisher Scientific) aiming for a minimum recovery of 15 ng sequencing library to load onto the flow cell.

Sequencing libraries were sequenced using R9.4.1 Flow Cells (FLO-MIN106D, ONT) on the Mk1C device (ONT) using MinKNOW v.21.02.2 with real-time high accuracy base-calling and stringent demultiplexing (minimum barcoding score = 60). Within the barcoding options, barcoding on both ends and mid-read barcodes were both switched on. Reads were demultiplexed and binned in a barcode specific folder only if a barcode above the minimum barcoding score was identified on both read ends and if mid-read barcodes were not identified. Sequencing runs were stopped after around 24 h, and base-calling was allowed to finish before data handling.

Bioinformatics data analysis was performed as per the ARTIC protocol using a combination of the ARTIC EBOV (<https://artic.network/ebov/ebov-bioinformatics-sop.html>) and ARTIC SARS-CoV-2 (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>) pipelines. A few minor modifications to the ARTIC bioinformatics protocol were incorporated. The two initial steps described, base-calling with Guppy and demultiplexing, were omitted as these were both done on the Mk1C device in real-time during the sequencing run; subsequently, the bioinformatics analysis was initiated from the read-filtering step (ARTIC Guppyplex). In brief, the ARTIC Guppyplex program was used to collect reads for each barcode into a single fastq file, in the presence of a length filter to remove chimeric reads. Reads were filtered based on length with a minimum (option: --min-length) and maximum (option: --max-length) length cut-off based on the amplicon size used (For V3, V4 and V5 primer pools: --min-length 400 and --max-length 700, for Zaire-PHE primer pool: --min-length 350 and --max-length 650). The quality check was omitted because only reads with a quality score of greater than 7 were processed. After merging and filtering, the ARTIC MinION pipeline was used to obtain the consensus sequences. The data were normalized to 200 and, using the --scheme-directory option, the pipeline was directed to the respective primer scheme used for each barcode. Reads were aligned to the NCBI reference KJ660347 (*Zaire ebolavirus* isolate H.sapiens-wt/GIN/2014/Makona-Gueckedou-C07) for data generated using V3, V4, and V5 primer pools and to NC\_002549.1 (*Zaire ebolavirus* isolate Ebola virus/H.sapiens-tc/COD/1976/Yambuku-Mayinga) for data generated using Zaire-PHE primer pools.

## Sequencing at IPD

Viral RNA was extracted from 140 µl of whole blood samples using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions and eluted in nuclease-free water to a final volume of 60 µl. Extracted RNA was tested using qRT-PCR as previously described<sup>46</sup>. In brief, the DNA library was prepared and enriched using the Illumina RNA Prep with Enrichment (L) Tagmentation kit (Illumina) according to the manufacturer's recommendations with a pan viral probe panel that included EBOV-specific targets<sup>5</sup>. The purified libraries were pooled and sequenced on the Illumina MiSeq platform using the MiSeq Reagents Kit v3 (Illumina) according to

the manufacturer's instructions. Illumina sequence reads were quality trimmed by Prinseq-lite and consensus EBOV genome sequences were generated using an in-house de novo genome assembly pipeline.

## Phylogenetic analysis of full-length genome sequences

### Phylogenetic inference

The new EBOV genome sequences were embedded in different datasets for subsequent analyses. For phylogenetic reconstruction, we use a Zaire Ebola virus dataset consisting of 55 representative genomes from previous outbreaks and a Makona virus dataset consisting of 1,065 genomes sampled from Guinea, Sierra Leone and Liberia between 2014 and 2015. Multiple sequence alignment was performed using mafft<sup>47</sup>. We identified 6 T-to-C mutations in the genome from patient 11 that were indicative of mutations induced by adenosine deaminases acting on RNA. According to previous recommendations<sup>48</sup>, we masked these positions in this genome in all further analyses. Maximum likelihood trees were reconstructed using IQ-TREE under the GTR model with gamma (G) distributed rate variation among sites<sup>49</sup>. Temporal divergence plots of genetic divergence from the root of phylogenies against sampling time were constructed using TempEst<sup>50</sup>. To construct the temporal divergence plot for the Guinean 2021 genome data, we used a tree reconstructed under an HKY+G model.

### Local molecular-clock model analysis

We used BEAST to fit a local molecular-clock model to a dataset consisting of 1,020 dated Makona virus genomes and one of the 2021 genomes (patient 1)<sup>51,52</sup>. We specified a separate rate on the tip branch for this genome as well as on the tip branch for a genome in a 2016 outbreak. We used the Skygrid coalescent model as a flexible nonparametric tree prior and an HKY+G substitution model<sup>53</sup>.

### Guinea 2021 tMRCA estimation

Temporal signal was evaluated using the BETS procedure<sup>54</sup>. We estimated a slightly lower log marginal likelihood for a model that uses tip dates (-26,063.6) compared to a model that assumes sequences are sampled at the same time (-26,062.1). These BEAST analyses were performed using an exponential growth model, a strict molecular-clock model and an HKY+G substitution model. We specified a lognormal prior with a mean of 1 and a standard deviation of 5 on the population size and a Laplace prior with a scale of 100 on the growth rate. Default priors were used for all other parameters. For the estimation of divergence time, we used a normal prior on the

substitution rate with a mean of 0.001 and a standard deviation of 0.00004 based on the background EBOV rate estimated by the local molecular-clock analysis.

## Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

## Data availability

Sequencing results were made publicly available on 12 March 2021 through joint posting on <https://virological.org/c/ebolavirus/guinea-2021/44>. The sequences generated at CERFIG have been deposited to GitHub under project link [https://github.com/kabinet1980/Ebov\\_Guinea2021/blob/main/EBOV\\_Guinea\\_2021\\_genomes\\_CERFIG.fasta](https://github.com/kabinet1980/Ebov_Guinea2021/blob/main/EBOV_Guinea_2021_genomes_CERFIG.fasta) and at the European Nucleotide Archive (ENA) under accession code [PRJEB43650](#). The sequences generated at PFHVG have been deposited to GitHub under project link <https://github.com/PFHVG/EBOVsequencing> and the genome sequences for the two samples at IPD are available at <https://drive.google.com/drive/folders/14dfGdNjWw17TkjrEQKLCrwlJ4WBBHI6K>. Genome sequences are also available at the NCBI GenBank under accession codes [ERX5245591](#) to [ERX5245598](#); [MZ424849](#) to [MZ424862](#); [MZ605320](#) and [MZ605321](#).

## Code availability

All the codes for the analyses presented in this paper, including the analysis pipeline, is described in detail in Methods and is available in published papers and public websites or, for in-house pipelines, is available upon reasonable request from the corresponding author.

## References

1. 1.

Baize, S. et al. Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014).

2. 2.

World Health Organization. *2016 Situation Report: Ebola Virus Disease* (World Health Organization, 2016).

3. 3.

Malvy, D., McElroy, A. K., de Clerck, H., Günther, S. & van Griensven, J. Ebola virus disease. *Lancet* **393**, 936–948 (2019).

4. 4.

Mbala-Kingebeni, P. et al. Rapid confirmation of the Zaire Ebola virus in the outbreak of the Equateur Province in the Democratic Republic of Congo: implications for public health interventions. *Clin. Infect. Dis.* **68**, 330–333 (2019).

5. 5.

Mbala-Kingebeni, P. et al. 2018 Ebola virus disease outbreak in Equateur Province, Democratic Republic of the Congo: a retrospective genomic characterization. *Lancet Infect. Dis.* **19**, 641–647 (2019).

6. 6.

Pigott, D. M. et al. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**, e04395 (2014).

7. 7.

Diallo, B. et al. Resurgence of Ebola virus disease in Guinea linked to a survivor with virus persistence in seminal fluid for more than 500 days. *Clin. Infect. Dis.* **63**, 1353–1356 (2016).

8. 8.

Sissoko, D. et al. Ebola virus persistence in breast milk after no reported illness: a likely source of virus transmission from mother to child. *Clin. Infect. Dis.* **64**, 513–516 (2017).

9. 9.

Keita, A. K. et al. A 40-month follow-up of Ebola virus disease survivors in Guinea (PostEbogui) reveals long-term detection of Ebola viral ribonucleic acid in semen and breast milk. *Open Forum Infect. Dis.* **6**, ofz482 (2019).

10. 10.

Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).

11. 11.  
Diehl, W. E. et al. Ebola virus glycoprotein with increased infectivity dominated the 2013–2016 epidemic. *Cell* **167**, 1088–1098.e6 (2016).
12. 12.  
Urbanowicz, R. A. et al. Human adaptation of Ebola virus during the West African outbreak. *Cell* **167**, 1079–1087.e5 (2016).
13. 13.  
Blackley, D. J. et al. Reduced evolutionary rate in reemerged Ebola virus transmission chains. *Sci. Adv.* **2**, e1600378 (2016).
14. 14.  
Diallo, M. S. K. et al. Understanding the long-term evolution and predictors of sequelae of Ebola virus disease survivors in Guinea: a 48-month prospective, longitudinal cohort study (PostEboGui). *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab168> (2021).
15. 15.  
Etard, J. F. et al. Multidisciplinary assessment of post-Ebola sequelae in Guinea (PostEboGui): an observational cohort study. *Lancet Infect. Dis.* **17**, 545–552 (2017).
16. 16.  
PREVAIL III Study Group. A longitudinal study of Ebola sequelae in Liberia. *N. Engl. J. Med.* **380**, 924–934 (2019).
17. 17.  
Diallo, M. S. K. et al. Prevalence of infection among asymptomatic and paucisymptomatic contact persons exposed to Ebola virus in Guinea: a retrospective, cross-sectional observational study. *Lancet Infect. Dis.* **19**, 308–316 (2019).
18. 18.  
Glynn, J. R. et al. Asymptomatic infection and unrecognised Ebola virus disease in Ebola-affected households in Sierra Leone: a cross-sectional study using a new

non-invasive assay for antibodies to Ebola virus. *Lancet Infect. Dis.* **17**, 645–653 (2017).

19. 19.

Camara, I. et al. Unrecognized Ebola virus infection in Guinea: complexity of surveillance in a health crisis situation: case report. *Pan Afr. Med. J.* **36**, 201 (2020).

20. 20.

Thorson, A. E. et al. Persistence of Ebola virus in semen among Ebola virus disease survivors in Sierra Leone: a cohort study of frequency, duration, and risk factors. *PLoS Med.* **18**, e1003273 (2021).

21. 21.

Fischer, W. A. et al. Ebola virus ribonucleic acid detection in semen more than two years after resolution of acute Ebola virus infection. *Open Forum Infect. Dis.* **4**, ofx155 (2017).

22. 22.

Sissoko, D. et al. Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. *Lancet Glob. Health* **5**, e80–e88 (2017).

23. 23.

Mate, S. E. et al. Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* **373**, 2448–2454 (2015).

24. 24.

Dokubo, E. K. et al. Persistence of Ebola virus after the end of widespread transmission in Liberia: an outbreak report. *Lancet Infect. Dis.* **18**, 1015–1024 (2018).

25. 25.

Liu, W. J. et al. Comprehensive clinical and laboratory follow-up of a female patient with Ebola virus disease: Sierra Leone Ebola virus persistence study. *Open Forum Infect. Dis.* **6**, ofz068 (2019).

26. 26.

Wiedemann, A. et al. Long-lasting severe immune dysfunction in Ebola virus disease survivors. *Nat. Commun.* **11**, 3730 (2020).

27. 27.

Adaken, C. et al. Ebola virus antibody decay–stimulation in a high proportion of survivors. *Nature* **590**, 468–472 (2021).

28. 28.

Thom, R. et al. Longitudinal antibody and T cell responses in Ebola virus disease survivors and contacts: an observational cohort study. *Lancet Infect. Dis.* **21**, 507–516 (2021).

29. 29.

Mbala Kingebeni, P. et al. Ebola virus transmission initiated by relapse of systemic Ebola virus disease. *N. Engl. J. Med.* **384**, 1240–1247 (2021).

30. 30.

Muyembe-Tamfum, J. J. et al. Two Ebola virus variants circulating during the 2020 Equateur Province outbreak. <https://virological.org/t/two-ebola-virus-variants-circulating-during-the-2020-equateur-province-outbreak/538> (2020).

31. 31.

Leroy, E. M. et al. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Dis.* **9**, 723–728 (2009).

32. 32.

Grard, G. et al. Emergence of divergent Zaire Ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *J. Infect. Dis.* **204** (Suppl 3), S776–S784 (2011).

33. 33.

Enria, L. What Crisis Produces: Dangerous Bodies, Ebola Heroes and Resistance in Sierra Leone' *BathPapers in International Development and Wellbeing*, Centre for Development Studies, University of Bath no. 53 (2017).

34. 34.

Ronse, M. et al. What motivates Ebola survivors to donate plasma during an emergency clinical trial? The case of Ebola-Tx in Guinea. *PLoS Negl. Trop. Dis.* **12**, e0006885 (2018).

35. 35.

Sow, S., Desclaux, A., Taverne, B. & Groupe d'étude PostEboGui. Ebola en Guinée: formes de la stigmatisation des acteurs de santé survivants. *Bull. Soc. Pathol. Exot.* **109**, 309–313 (2016).

36. 36.

Farmer, P. *Aids and Accusation. Haiti and the Geography of Blame* (Univ. California Press, 2006).

37. 37.

Wilkinson, A. & Fairhead, J. Comparison of social resistance to Ebola response in Sierra Leone and Guinea suggests explanations lie in political configurations not culture. *Crit. Public Health* **27**, 14–27 (2017).

38. 38.

Gire, S. K. et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).

39. 39.

Arias, A. et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol.* **2**, vew016 (2016).

40. 40.

Pini, A. et al. Field investigation with real-time virus genetic characterisation support of a cluster of Ebola virus disease cases in Dubréka, Guinea, April to June 2015. *Euro Surveill.* **23**, 17-00140 (2018).

41. 41.

Mbala-Kingebeni, P. et al. Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri Provinces of the Democratic

Republic of the Congo: a rapid genomic assessment. *Lancet Infect. Dis.* **19**, 648–657 (2019).

42. 42.

Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

43. 43.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

44. 44.

Bolger A. M., Lohse M. & Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

45. 45.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

46. 46.

Weidmann, M., Mühlberger, E. & Hufert, F. T. Rapid detection protocol for filoviruses. *J. Clin. Virol.* **30**, 94–99 (2004).

47. 47.

Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).

48. 48.

Dudas, G. et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315 (2017).

49. 49.

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).

50. 50.

Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

51. 51.

Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).

52. 52.

Worobey, M., Han, G. Z. & Rambaut, A. A. Synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* **508**, 254–257 (2014).

53. 53.

Gill, M. S. et al. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).

54. 54.

Duchene, S., et al. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* **6**, veaa061 (2020).

## Acknowledgements

We thank the ANSS and the Ministry of Health of the Republic of Guinea, the healthcare workers (medical doctors, nurses and laboratory technicians) from the treatment centres in Nzérékoré and Nongo, Conakry; the laboratory personnel in Guékédou, INSP (Conakry) and CERFIG (Conakry). CERFIG also acknowledges J. Gogarten and bioinformatics support from RKI. The UK Health Security Agency would like to thank Oxford Nanopore Technologies for the donation of reagents and equipment to support the setting up of the sequencing capacity at PFHG. BNITM thanks the ARTIC Network (<https://artic.network/>). The work of CERFIG and TransVIHMI was supported in part by grants from the EBO-SURSY Project funded by the European Union, International Mixt Laboratory ‘RESPIRE’ of IRD (Institut de Recherche pour le Développement), Montpellier Université d’Excellence (EBOHEALTH; I-Site MUSE, ANR-16-IDEX-0006) and Institut National de la Santé et de la Recherche Médicale (INSERM)/REACTing (REsearch and ACTion targeting

emerging infectious diseases). The work of the RKI was partly funded by the German Ministry of Health ‘Global Protection Program’ project TRICE. F.L.M. received funding through the program ‘EBOVAC3 Bringing a prophylactic Ebola vaccine to licensure’ funded by Innovative Medicine Initiative (grant agreement number 800176) and run by London School of Hygiene and Tropical Medicine and INSERM. The work of PFHG and BNITM was supported by the German Federal Ministry of Health through support of the WHO Collaborating Centre for Arbovirus and Hemorrhagic Fever Viruses at the BNITM (agreement ZMV I1-2517WHO005), and through the Global Health Protection Programme (GHPP, agreements ZMV I1-2517GHP-704 and ZMVI1-2519GHP704), and by the Coalition for Epidemic Preparedness Innovations (CEPI). The work of BNITM and the UK Health Security Agency was further supported by the Research and Innovation Programme of the European Union under Horizon 2020 grant agreement no. 871029-EVA-GLOBAL. The European Mobile Laboratory (EMLab) coordinated by BNITM is a technical partner of the WHO Global Outbreak Alert and Response Network (GOARN) and the deployment of EMLab experts and sequencing capacities to Guinea was coordinated and supported by the GOARN Operational Support Team at WHO/HQ and the WHO country office in Guinea. The research leading to these results has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 725422-ReservoirDOCS). The ARTIC Network receives funding from the Wellcome Trust through project 206298/Z/17/Z. P.L. acknowledges support by the Research Foundation – Flanders (‘Fonds voor Wetenschappelijk Onderzoek – Vlaanderen’, G066215N, G0D5117N and G0B9317N).

## Author information

### Author notes

1. These authors contributed equally: Alpha K. Keita, Fara R. Koundouno, Martin Faye, Ariane Düx, Julia Hinzmann
2. These authors jointly supervised this work: Sophie Duraffour, Ousmane Faye, Fabian Leendertz, Martine Peeters, Abdoulaye Toure, N’Faly Magassouba

### Affiliations

1. Centre de Recherche et de Formation en Infectiologie de Guinée (CERFIG), Université de Conakry, Conakry, Guinea  
Alpha Kabinet Keita, Haby Diallo, Frederic Le Marcis, Mamadou S. Sow, Abdoul K. Soumah, Ibrahima Camara, Moriba Povogui & Abdoulaye Toure

2. TransVIHMI, Montpellier University/IRD/INSERM, Montpellier, France

Alpha Kabinet Keita, Ahidjo Ayouba, Frederic Le Marcis, Nicole Vidal, Eric Delaporte & Martine Peeters

3. Laboratoire du Projet des Fièvres Hémorragiques de Guinée (PFHG), Conakry, Guinea

Fara R. Koundouno, Barré Soropogui, Kékoura Ifono, Joseph A. Bore, Jacob Camara, Saa L. Millimono, Ahmadou Doré, Karifa Kourouma, Assaïtou Bah & N'. Faly Magassouba

4. Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany

Fara R. Koundouno, Julia Hinzmann, Kékoura Ifono, Annick Renevey, Saa L. Millimono, Nathalie J. Vielle, Karifa Kourouma, Giuditta Annibaldis, Anke Thielebein, Meike Pahlmann, Stephan Günther, Liana E. Kafetzopoulou & Sophie Duraffour

5. Institut Pasteur de Dakar (IPD), Dakar, Senegal

Martin Faye, Moussa M. Diagne, Amadou Diallo, Mamadou Diop, Cheikh Loucoubar, Christophe Peyrefitte, Amadou A. Sall & Ousmane Faye

6. Robert Koch Institute (RKI), Berlin, Germany

Ariane Düx, Sébastien Calvignac-Spencer, Almudena Mari-Saez & Fabian H. Leendertz

7. German Center for Infection Research (DZIF), Partner Site Hamburg–Lübeck–Borstel–Riems, Hamburg, Germany

Julia Hinzmann, Annick Renevey, Anke Thielebein, Meike Pahlmann, Stephan Günther, Liana E. Kafetzopoulou & Sophie Duraffour

8. UK Health Security Agency, National Infection Service, Porton Down, Salisbury, UK

Julia Hinzmann, Steven T. Pullan & Miles W. Carroll

9. Ecole Nationale Supérieure de Lyon, Lyon, France

Frédéric Le Marcis & Amadou Sidibe

10. Hôpital National Donka, Service des Maladies Infectieuses et Tropicales, Conakry, Guinea

Mamadou S. Sow, Fodé Y. Soumah, Moussa Baldé, Dembo Diakite, Fodé B. Sako & Fodé A. Traore
11. Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Joseph A. Bore, Steven T. Pullan & Miles W. Carroll
12. Institut National de Santé Publique de Guinée (INSP), Conakry, Guinea

Mamadou B. Keita, Kaka Kourouma & Abdoulaye Toure
13. World Health Organization (WHO), Conakry, Guinea

Nathalie J. Vielle, Giuditta Annibaldi, Georges A. Ki-Zerbo, Stephan Günther & Sophie Duraffour
14. Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK

Joshua Quick
15. Health Emergencies Program, World Health Organization (WHO), Geneva, Switzerland

Pierre Formenty & Anais Legand
16. PraesensBio, Lincoln, NE, USA

Karla Pietro & Michael R. Wiley
17. University of Nebraska Medical Center, Omaha, NE, USA

Michael R. Wiley
18. Institut Pasteur de Guinée, Conakry, Guinea

Noel Tordo
19. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

John T. McCrone & Andrew Rambaut

20. Hôpital Régional de Nzérékoré, Nzérékoré, Guinea

Youssouf Sidibé, Mamadou D. Barry, Madeleine Kourouma, Cé D.  
Saouromou & Mamadou Condé

21. Agence Nationale de Sécurité Sanitaire, Conakry, Guinea

Sakoba Keita

22. Direction Nationale des Laboratoires, Ministère de la Santé, Conakry, Guinea

Mandiou Diakite & Mamadou S. Bah

23. Université Gamal Abdel Nasser de Conakry, Conakry, Guinea

Mandiou Diakite

24. Department of Microbiology, Immunology and Transplantation, Rega Institute,  
KU Leuven, Leuven, Belgium

Philippe Lemey & Liana E. Kafetzopoulou

25. Infectious Diseases Departement, University Hospital Montpellier, Montpellier,  
France

Eric Delaporte

## Contributions

A.A.S., A.K.K., A. Toure, E.D., F.H.L., F.R.K., L.E.K., M. Peeters, N.'F.M., O.F., S.D. and S.G. conceived and designed the study. A.B., A. Doré, A.K.K., A.K.S., A.M.-S., A.S., B.S., C.D.S., D.D., F.L., F.L.M. F.R.K., F.Y.S., F.A.T., F.B.S., G.A., H.D., I.C., J.A.B., J.C., K.I., Kaka Kourouma, Karifa Kourouma, S.L.M., M.B., M.B.K., M.C., M.D.B., M.K., M. Povogui, M.S.S., N.V., N.J.V., S.D. and Y.S. collected data and/or performed medical examinations and/or laboratory diagnostics. A.A., A. Düx, A. Renevey, B.S., H.D., J.A.B., J.H., K.I., K.P., M.F., M.M.D. and S.C.-S. performed sequencing and/or sequence validation. A.A., A. Rambaut, J.H., J.T.M., L.E.K., M.R.W., P.L., S.C.-S. and S.D. performed formal phylogenetic analysis. A. Diallo, C.L., F.L.M. and M. Diop performed data analysis. E.D., F.H.L., M. Peeters, M.W.C., S.D. and S.G. acquired funding. F.H.L., K.P., M.R.W. and S.C.-S. provided reagents. A.K.K., A.L., A. Thielebein, A. Toure, C.P., E.D., F.H.L., G.A.K.-Z., J.Q., M. Diakite, M. Pahlmann, M. Peeters, M.S.B., M.W.C., N.'F.M., N.T., P.F., S.D., S.G., S.K. and S.T.P. implemented the project. A. Düx, A.K.K., M. Peeters and S.C.-S. wrote the first draft of the manuscript. A.A.S., A. Düx, A.K.K., A.M.-S., A. Renevey, A. Rambaut,

A. Toure, E.D., F.H.L., F.L., F.L.M. L.E.K., M.F., M.M.D., M. Peeters, N.'F.M., O.F., P.L., S.C.-S., S.D. and S.G. wrote and edited the manuscript. All authors read and approved the contents of the manuscript.

## Corresponding author

Correspondence to Alpha Kabinet Keita.

## Ethics declarations

### Competing interests

M.W.C. received materials for this study from Oxford Nanopore Technologies. All other authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Robert Garry, Joel Montgomery and Michael Worobey for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

### [Supplementary Information](#)

This file contains Supplementary Figs. 1 and 2, and Supplementary Table 1.

### [Reporting Summary](#)

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### [Cite this article](#)

Keita, A.K., Koundouno, F.R., Faye, M. *et al.* Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. *Nature* **597**, 539–543 (2021).  
<https://doi.org/10.1038/s41586-021-03901-9>

- Received: 06 April 2021
- Accepted: 11 August 2021
- Published: 15 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03901-9>

## 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

## [\*\*Ebola virus can lie low and reactivate after years in human survivors\*\*](#)

- Robert F. Garry

News & Views 15 Sept 2021

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03901-9>

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

- Article
- [Published: 15 September 2021](#)

# An engineered IL-2 partial agonist promotes CD8<sup>+</sup> T cell stemness

- [Fei Mo](#) ORCID: [orcid.org/0000-0002-8601-2568](https://orcid.org/0000-0002-8601-2568)<sup>1,nal</sup>,
- [Zhiya Yu](#)<sup>2,nal</sup>,
- [Peng Li](#)<sup>1,nal</sup>,
- [Jangsuk Oh](#)<sup>1</sup>,
- [Rosanne Spolski](#)<sup>1</sup>,
- [Liang Zhao](#) ORCID: [orcid.org/0000-0002-2126-6778](https://orcid.org/0000-0002-2126-6778)<sup>3</sup>,
- [Caleb R. Glassman](#) ORCID: [orcid.org/0000-0002-3342-7989](https://orcid.org/0000-0002-3342-7989)<sup>4</sup>,
- [Tori N. Yamamoto](#) ORCID: [orcid.org/0000-0003-1965-7822](https://orcid.org/0000-0003-1965-7822)<sup>2</sup>,
- [Yun Chen](#)<sup>6</sup>,
- [Filip M. Golebiowski](#) ORCID: [orcid.org/0000-0002-8011-2162](https://orcid.org/0000-0002-8011-2162)<sup>6</sup>,
- [Dalton Hermans](#)<sup>1</sup>,
- [Sonia Majri-Morrison](#)<sup>4</sup>,
- [Lora K. Picton](#)<sup>4,5</sup>,
- [Wei Liao](#)<sup>1</sup>,
- [Min Ren](#)<sup>1</sup>,
- [Xiaoxuan Zhuang](#)<sup>7</sup>,
- [Suman Mitra](#)<sup>1</sup>,
- [Jian-Xin Lin](#)<sup>1</sup>,
- [Luca Gattinoni](#) ORCID: [orcid.org/0000-0003-2239-3282](https://orcid.org/0000-0003-2239-3282)<sup>8,9,10</sup>,
- [Jonathan D. Powell](#)<sup>3</sup>,
- [Nicholas P. Restifo](#) ORCID: [orcid.org/0000-0003-4229-4580](https://orcid.org/0000-0003-4229-4580)<sup>2</sup>,
- [K. Christopher Garcia](#) ORCID: [orcid.org/0000-0001-9273-0278](https://orcid.org/0000-0001-9273-0278)<sup>4,5</sup>  
&
- [Warren J. Leonard](#) ORCID: [orcid.org/0000-0002-5740-7448](https://orcid.org/0000-0002-5740-7448)<sup>1</sup>

[Nature](#) volume 597, pages 544–548 (2021)

- 8704 Accesses
- 32 Altmetric
- [Metrics details](#)

## Subjects

- [Cytokines](#)
- [Tumour immunology](#)

## Abstract

Adoptive transfer of antigen-specific T cells represents a major advance in cancer immunotherapy, with robust clinical outcomes in some patients<sup>1</sup>. Both the number of transferred T cells and their differentiation state are critical determinants of effective responses<sup>2,3</sup>. T cells can be expanded with T cell receptor (TCR)-mediated stimulation and interleukin-2, but this can lead to differentiation into effector T cells<sup>4,5</sup> and lower therapeutic efficacy<sup>6</sup>, whereas maintenance of a more stem-cell-like state before adoptive transfer is beneficial<sup>7</sup>. Here we show that H9T, an engineered interleukin-2 partial agonist, promotes the expansion of CD8<sup>+</sup> T cells without driving terminal differentiation. H9T led to altered STAT5 signalling and mediated distinctive downstream transcriptional, epigenetic and metabolic programs. In addition, H9T treatment sustained the expression of T cell transcription factor 1 (TCF-1) and promoted mitochondrial fitness, thereby facilitating the maintenance of a stem-cell-like state. Moreover, TCR-transgenic and chimeric antigen receptor-modified CD8<sup>+</sup> T cells that were expanded with H9T showed robust anti-tumour activity *in vivo* in mouse models of melanoma and acute lymphoblastic leukaemia. Thus, engineering cytokine variants with distinctive properties is a promising strategy for creating new molecules with translational potential.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

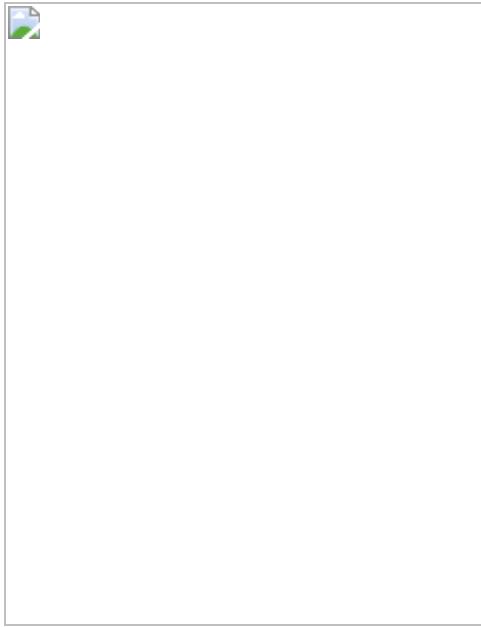
[Rent or Buy](#)

All prices are NET prices.

### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

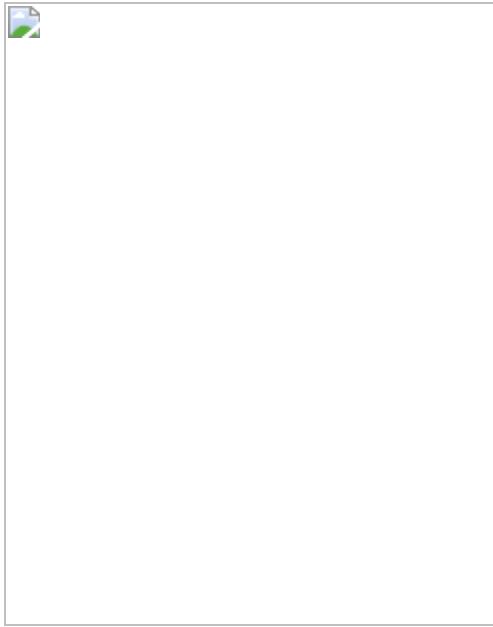
**Fig. 1: Differential effects of H9T versus IL-2 and H9 on CD8<sup>+</sup> T cells.**



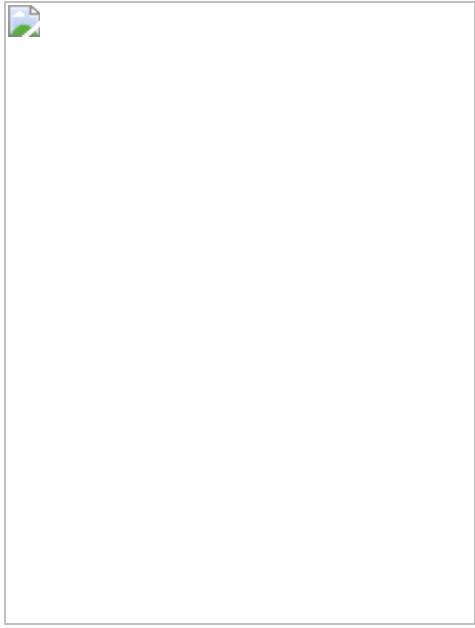
**Fig. 2: Transcriptional profile and epigenetic landscape of H9T-expanded CD8<sup>+</sup> T cells.**



**Fig. 3: Altered metabolism in H9T-expanded CD8<sup>+</sup> T cells.**



**Fig. 4: Increased anti-tumour activity of H9T-expanded CD8<sup>+</sup> T cells.**



## Data availability

Our ATAC-seq, ChIP-seq and RNA-seq data are available at the NCBI Gene Expression Omnibus (GEO) under the accession number GSE138698. Publicly available previously generated ChIP-seq<sup>56</sup> (GSE36890) and ATAC-seq<sup>22</sup> (GSE88987) data were also used in this study. [Source data](#) are provided with this paper.

# References

1. 1.

Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* 348, 62–68 (2015).

2. 2.

Shah, N. N. & Fry, T. J. Mechanisms of resistance to CAR T cell therapy. *Nat. Rev. Clin. Oncol.* 16, 372–385 (2019).

3. 3.

Crompton, J. G., Sukumar, M. & Restifo, N. P. Uncoupling T-cell expansion from effector differentiation in cell-based immunotherapy. *Immunol. Rev.* 257, 264–276 (2014).

4. 4.

Pipkin, M. E. et al. Interleukin-2 and inflammation induce distinct transcriptional programs that promote the differentiation of effector cytolytic T cells. *Immunity* 32, 79–90 (2010).

5. 5.

Kalia, V. et al. Prolonged interleukin-2R $\alpha$  expression on virus-specific CD8 $^{+}$  T cells favors terminal-effector differentiation *in vivo*. *Immunity* 32, 91–103 (2010).

6. 6.

Gattinoni, L., Speiser, D. E., Lichtenfeld, M. & Bonini, C. T memory stem cells in health and disease. *Nat. Med.* 23, 18–27 (2017).

7. 7.

Gattinoni, L., Klebanoff, C. A. & Restifo, N. P. Paths to stemness: building the ultimate antitumour T cell. *Nat. Rev. Cancer* 12, 671–684 (2012).

8. 8.

Silva, D. A. et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* 565, 186–191 (2019).

9. 9.

Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nat. Rev. Immunol.* 18, 648–659 (2018).

10. 10.

Leonard, W. J., Lin, J. X. & O’Shea, J. J. The  $\gamma_c$  family of cytokines: basic biology to therapeutic ramifications. *Immunity* 50, 832–850 (2019).

11. 11.

Kinter, A. L. et al. The common  $\gamma$ -chain cytokines IL-2, IL-7, IL-15, and IL-21 induce the expression of programmed death-1 and its ligands. *J. Immunol.* 181, 6738–6746 (2008).

12. 12.

Mujib, S. et al. Antigen-independent induction of Tim-3 expression on human T cells by the common  $\gamma$ -chain cytokines IL-2, IL-7, IL-15, and IL-21 is associated with proliferation and is dependent on the phosphoinositide 3-kinase pathway. *J. Immunol.* 188, 3745–3756 (2012).

13. 13.

Jin, H. T. et al. Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. *Proc. Natl Acad. Sci. USA* 107, 14733–14738 (2010).

14. 14.

Sakuishi, K. et al. Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J. Exp. Med.* 207, 2187–2194 (2010).

15. 15.

Levin, A. M. et al. Exploiting a natural conformational switch to engineer an interleukin-2 ‘superkine’. *Nature* 484, 529–533 (2012).

16. 16.

Mitra, S. et al. Interleukin-2 activity can be fine tuned with engineered receptor signaling clamps. *Immunity* 42, 826–838 (2015).

17. 17.

Wang, X., Rickert, M. & Garcia, K. C. Structure of the quaternary complex of interleukin-2 with its  $\alpha$ ,  $\beta$ , and  $\gamma_c$  receptors. *Science* 310, 1159–1163 (2005).

18. 18.

Liu, D. V., Maier, L. M., Hafler, D. A. & Wittrup, K. D. Engineered interleukin-2 antagonists for the inhibition of regulatory T cells. *J. Immunother.* 32, 887–894 (2009).

19. 19.

Gattinoni, L. et al. A human memory T cell subset with stem cell-like properties. *Nat. Med.* 17, 1290–1297 (2011).

20. 20.

McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu. Rev. Immunol.* 37, 457–495 (2019).

21. 21.

Im, S. J. et al. Defining CD8<sup>+</sup> T cells that provide the proliferative burst after PD-1 therapy. *Nature* 537, 417–421 (2016).

22. 22.

Scott-Browne, J. P. et al. Dynamic changes in chromatin accessibility occur in CD8<sup>+</sup> T cells responding to viral infection. *Immunity* 45, 1327–1340 (2016).

23. 23.

Wang, R. & Green, D. R. Metabolic checkpoints in activated T cells. *Nat. Immunol.* 13, 907–915 (2012).

24. 24.

Sukumar, M. et al. Inhibiting glycolytic metabolism enhances CD8<sup>+</sup> T cell memory and antitumor function. *J. Clin. Invest.* 123, 4479–4488 (2013).

25. 25.

van der Windt, G. J. et al. Mitochondrial respiratory capacity is a critical regulator of CD8<sup>+</sup> T cell memory development. *Immunity* 36, 68–78 (2012).

26. 26.

Manjunath, N. et al. Effector differentiation is not prerequisite for generation of memory cytotoxic T lymphocytes. *J. Clin. Invest.* 108, 871–878 (2001).

27. 27.

Hinrichs, C. S. et al. IL-2 and IL-21 confer opposing differentiation programs to CD8<sup>+</sup> T cells for adoptive immunotherapy. *Blood* 111, 5326–5333 (2008).

28. 28.

Onishi, M. et al. Identification and characterization of a constitutively active STAT5 mutant that promotes cell proliferation. *Mol. Cell Biol.* 18, 3871–3879 (1998).

29. 29.

Blank, C. U. et al. Defining ‘T cell exhaustion’. *Nat. Rev. Immunol.* 19, 665–674 (2019).

30. 30.

Yang, Q. et al. TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat. Immunol.* 16, 1044–1050 (2015).

31. 31.

Liu, X. et al. Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* 11, 179–186 (1997).

32. 32.

Udy, G. B. et al. Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression. *Proc. Natl Acad. Sci. USA* 94, 7239–7244 (1997).

33. 33.

Vodnala, S. K. et al. T cell stemness and dysfunction in tumors are triggered by a common mechanism. *Science* 363, eaau0135 (2019).

34. 34.

Hanada, K. I., Yu, Z., Chappell, G. R., Park, A. S. & Restifo, N. P. An effective mouse model for adoptive cancer immunotherapy targeting neoantigens. *JCI Insight* 4, e124405 (2019).

35. 35.

Zhu, E. F. et al. Synergistic innate and adaptive immune response to combination immunotherapy with anti-tumor antigen antibodies and extended serum half-life IL-2. *Cancer Cell* 27, 489–501 (2015).

36. 36.

Sockolosky, J. T. et al. Selective targeting of engineered T cells using orthogonal IL-2 cytokine–receptor complexes. *Science* 359, 1037–1042 (2018).

37. 37.

Bijl, J., Sauvageau, M., Thompson, A. & Sauvageau, G. High incidence of proviral integrations in the Hoxa locus in a new model of E2a–PBX1-induced B-cell leukemia. *Genes Dev.* 19, 224–233 (2005).

38. 38.

Jacoby, E. et al. Murine allogeneic CD19 CAR T cells harbor potent antileukemic activity but have the potential to mediate lethal GVHD. *Blood* 127, 1361–1370 (2016).

39. 39.

Yamamoto, T. N. et al. T cells genetically engineered to overcome death signaling enhance adoptive cancer immunotherapy. *J. Clin. Invest.* 129, 1551–1565 (2019).

40. 40.

Ya, Z., Hailemichael, Y., Overwijk, W. & Restifo, N. P. Mouse model for pre-clinical study of human cancer immunotherapy. *Curr. Protoc. Immunol.* 108, 20.21.21–20.21.43 (2015).

41. 41.

Kochenderfer, J. N., Yu, Z., Frasher, D., Restifo, N. P. & Rosenberg, S. A. Adoptive transfer of syngeneic T cells transduced with a chimeric antigen receptor that recognizes murine CD19 can eradicate lymphoma and normal B cells. *Blood* 116, 3875–3886 (2010).

42. 42.

van der Windt, G. J., Chang, C. H. & Pearce, E. L. Measuring bioenergetics in T cells using a seahorse extracellular flux analyzer. *Curr. Protoc. Immunol.* 113, 3.16B.1–13.16B.14 (2016).

43. 43.

Hermans, D. et al. Lactate dehydrogenase inhibition synergizes with IL-21 to promote CD8<sup>+</sup> T cell stemness and antitumor immunity. *Proc. Natl Acad. Sci. USA* 117, 6047–6055 (2020).

44. 44.

Sukumar, M. et al. Mitochondrial membrane potential identifies cells with enhanced stemness for cellular therapy. *Cell Metab.* 23, 63–76 (2016).

45. 45.

Lin, J. X. et al. Critical functions for STAT5 tetramers in the maturation and survival of natural killer cells. *Nat. Commun.* 8, 1320 (2017).

46. 46.

Li, P. et al. BATF–JUN is critical for IRF4-mediated transcription in T cells. *Nature* 490, 543–546 (2012).

47. 47.

Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.21–21.29.29 (2015).

48. 48.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009).

49. 49.

Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* 25, 1105–1111 (2009).

50. 50.

Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 (2011).

51. 51.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).

52. 52.

Wherry, E. J. et al. Molecular signature of CD8<sup>+</sup> T cell exhaustion during chronic viral infection. *Immunity* 27, 670–684 (2007).

53. 53.

Doering, T. A. et al. Network analysis reveals centrally connected genes and pathways involved in CD8<sup>+</sup> T cell exhaustion versus memory. *Immunity* 37, 1130–1144 (2012).

54. 54.

Wu, T. et al. The TCF1–Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci. Immunol.* 1, eaai8593 (2016).

55. 55.

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).

56. 56.

Lin, J. X. et al. Critical role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function. *Immunity* 36, 586–599 (2012).

## Acknowledgements

We thank R. Ahmed for critical comments; W. Yang and C. Wu for discussions; and the NHLBI DIR Flow Cytometry Core, DNA Sequencing and Genomics Core for assistance with cell sorting and DNA sequencing. The work was supported by the Division of Intramural Research at the National Heart, Lung, and Blood Institute, and by the Division of Intramural Research at the National Cancer Institute (NCI). K.C.G. was supported by National Institutes of Health (NIH) grant AI51321, the Parker Institute of Cancer Immunotherapy, the Ludwig Foundation, the Mathers Foundation and the Howard Hughes Medical Institute. J.D.P was supported by the NIH (R01AI07761, P41EB028239-01) and The Bloomberg–Kimmel Institute for Cancer Immunotherapy. N.P.R was supported by the Intramural Research Program of the NCI and the Cancer Moonshot Program for the Center for Cell-Based Therapy at the NCI, NIH. L.G. was supported by the Intramural Research Program of the US NIH, NCI, Center for Cancer Research (ZIA-BC011480).

## Author information

Author notes

1. These authors contributed equally: Fei Mo, Zhiya Yu, Peng Li

## Affiliations

1. Laboratory of Molecular Immunology and the Immunology Center, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Fei Mo, Peng Li, Jangsuk Oh, Rosanne Spolski, Dalton Hermans, Wei Liao, Min Ren, Suman Mitra, Jian-Xin Lin & Warren J. Leonard

2. Surgery Branch, National Cancer Institute, Bethesda, MD, USA

Zhiya Yu, Tori N. Yamamoto & Nicholas P. Restifo

3. Johns Hopkins University School of Medicine, Baltimore, MD, USA

Liang Zhao & Jonathan D. Powell

4. Department of Molecular and Cellular Physiology and Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA

Caleb R. Glassman, Sonia Majri-Morrison, Lora K. Picton & K. Christopher Garcia

5. Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA

Lora K. Picton & K. Christopher Garcia

6. National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

Yun Chen & Filip M. Golebiowski

7. Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD, USA

Xiaoxuan Zhuang

8. Center for Cancer Research, National Cancer Institute, Bethesda, MD,  
USA

Luca Gattinoni

9. Department of Functional Immune Cell Modulation, Regensburg  
Center for Interventional Immunology, Regensburg, Germany

Luca Gattinoni

10. University of Regensburg, Regensburg, Germany

Luca Gattinoni

## **Contributions**

F.M., Z.Y., J.O., L.Z., C.R.G., T.N.Y., D.H. and S.M. designed and performed experiments, and analysed data. F.M., J.O., Y.C., F.M.G., S.S.M., L.K.P. and M.R. purified protein. P.L. and F.M. analysed the bioinformatics data. R.S., W.L., X.Z., J.-X.L. and L.G. analysed data and edited the paper. J.D.P., N.P.R., K.C.G. and W.J.L. supervised the project and analysed data. F.M. and W.J.L. wrote the paper.

## **Corresponding authors**

Correspondence to Nicholas P. Restifo or K. Christopher Garcia or Warren J. Leonard.

## **Ethics declarations**

## **Competing interests**

W.J.L. K.C.G. and S.M. are inventors on patents and patent applications that include H9T. L.G. is an inventor on a patent that describes methods for the generation and isolation of stem-cell memory T ( $T_{scm}$ ) cells. L.G. has

consulting agreements with Lyell Immunopharma, AstraZeneca, Turnstone Biologics, Xcelcyte and Advaxis Immunotherapies. L.G. is on the scientific advisory board of Poseida Therapeutics and Kiromic, and is a stockholder of Poseida Therapeutics.

## Additional information

**Peer review information** *Nature* thanks Greg Delgoffe, Stephen Jameson and E. Wherry for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Characterization of H9T-expanded CD8<sup>+</sup> T cells.

**a**, Pre-activated mouse CD8<sup>+</sup> T cells were rested overnight and cultured with PBS, IL-2, H9, or H9T as indicated for 10 min, and western blotted for phospho-STAT5 (pSTAT5) and pERK. Total STAT5 and ERK were included as controls (on the same gel of pSTAT5 and pERK). Data are representative of two independent experiments. Relative densitometry is shown below each panel (normalized to the IL-2 condition). For gel source data, see Supplementary Fig. [1b](#). **b**, Pre-activated CD8<sup>+</sup> T cells were rested overnight and stimulated with 10 nM IL-2 or H9T, lysed, and immunoprecipitated with anti-STAT5A or anti-STAT5B antibodies followed with western blotting analysis for pSTAT5 and STAT5A or STAT5B as loading controls (on the same gel). Data are representative of two independent experiments. Relative densitometry is shown below each panel (normalized to the IL-2 condition). For gel source data, see Supplementary Fig. [1c](#). Data are from two independent repeats. **c**, Pre-activated CD8<sup>+</sup> T cells were rested overnight, cultured with PBS, IL-2, H9, or H9T for 6 days, fixed, permeabilized and intracellularly stained for pSTAT5;  $n = 4$  mice. Data are presented as mean values  $\pm$  SEM, one-way ANOVA test with

Dunnett's correction. Data are from two independent repeats. **d**, Pre-activated mouse CD8<sup>+</sup> T cells were rested overnight and then cultured in medium containing serial dilutions of IL-2, H9, or H9T for 8 days and analysed for TIM-3 expression. Data are representative of three independent experiments. **e–h**, Pre-activated CD8<sup>+</sup> T cells were cultured with 10 nM IL-2, H9, or H9T for 8 days, and surface expression of TIM-3 (**e**), PD-1 (**f**), LAG-3 (**g**), and 2B4 (**h**) was analysed by flow cytometry. Data are representative of four independent experiments

[Source data](#).

**Extended Data Fig. 2 Functional analysis of CD8<sup>+</sup> T cells expanded with IL-2, H9 or H9T.**

**a, b**, Expression of CD62L and the percentage of memory-like cells (CD62L<sup>+</sup>CD44<sup>+</sup>). Pre-activated mouse CD8<sup>+</sup> T cells were cultured with 10 nM IL-2, H9, or H9T for 8 days, and stained for CD44 and CD62L. Data are mean values  $\pm$  SEM,  $n = 6$  mice, one-way ANOVA test with Dunnett's correction. Data are representative of three independent experiments. **c**, Expression of CD62L in 8 day IL-2- or H9T-cultured cells or cells cultured for 6 days in H9T-containing medium and then switched to IL-2-containing medium for 2 days. Pre-activated CD8<sup>+</sup> T cells were cultured with 10 nM IL-2 or H9T, and on day 6, a fraction of the H9T-expanded cells was washed and subsequently cultured with 10 nM IL-2. Cells were collected two days later and analysed by flow cytometry. Data are presented as mean values  $\pm$  SEM,  $n = 4$  mice, one-way ANOVA test with Dunnett's correction. Data are representative of two independent experiments. **d–g**, Cytokine production and memory population in CD8<sup>+</sup> T cells expanded with IL-2, H9, or H9T. Pre-activated CD8<sup>+</sup> T cells were expanded for 8 days with IL-2, H9, or H9T, stimulated with 100 nM gp100 or control peptide for 1 h, and then treated with 5  $\mu$ g/ml Brefeldin A for 5 h. Cells were then fixed for intracellular staining of IFN- $\gamma$ ,  $n = 5$  mice (**d**); TNF,  $n = 5$  mice (**e**); IL-2,  $n = 3$  mice (**f**); and IL-10,  $n = 5$  mice (**g**); Data are presented as mean values  $\pm$  SEM, one-way ANOVA test with Dunnett's correction. Data are representative of two independent experiments

[Source data](#).

## **Extended Data Fig. 3 RNA-seq and ATAC-seq analysis of CD8<sup>+</sup> T cells expanded with IL-2, H9 or H9T.**

**a**, Volcano plots of RNA-seq data from pre-activated mouse CD8<sup>+</sup> T cells that were expanded for 6 days with IL-2, H9, or H9T. Shown are gene expression differences between cells expanded with IL-2 versus H9T (**a**) or H9 versus H9T (**b**). Data are representative of two independent experiments. **c**, Differentially expressed genes for CD8<sup>+</sup> T cells expanded for 6 days with H9 versus H9T. Data are representative of two independent experiments. **d**, GSEA analysis of RNA-seq data compared with endogenous memory versus exhausted populations of cells, with Kolmogorov–Smirnov test. Data are representative of two independent experiments. **e–g**, Pre-activated CD8<sup>+</sup> T cells were expanded for 8 days with IL-2, H9, or H9T and then either permeabilized for intracellular staining of TCF-1 (**e**) and BLIMP-1 (**f**) or analysed for surface expression of CXCR3 (**g**). Data are representative of three independent experiments. **h**, **i**, Expression of CD62L and TIM-3 in control or *Tcf7*-deficient CD8<sup>+</sup> T cells. CD8<sup>+</sup> T cells were isolated from *Tcf7* conditional knock-out mice (*Tcf7*<sup>−/−</sup>) or control littermates (*Tcf7*<sup>fl/fl</sup>) and activated for 2 days, rested overnight, followed with expansion in IL-2- or H9T-containing medium. Surface staining of CD62L (**h**) and TIM-3 (**i**) is shown. Data are mean ± SEM,  $n = 3$  mice, two-sided t-test. Data are representative of two independent experiments. **j**, Differentially expressed genes from RNA-seq were aligned to ATAC-seq plots. Data are representative of two independent experiments. **k**, **l**, ATAC-seq data from CD8<sup>+</sup> T cells expanded for 6 days were aligned with *in vivo* generated effector, exhausted, and memory populations; the ATAC-seq parts of **k** are also shown in Fig. 2j with ChIP-seq data. Shown is chromatin accessibility at the *Havcr2* (**k**) and *Il10* (**l**) loci. Data are representative of two independent experiments

[Source data](#).

## Extended Data Fig. 4 Metabolic profiling of IL-2-, H9- and H9T-expanded CD8<sup>+</sup> T cells.

**a**, Photograph showing the medium colour of mouse CD8<sup>+</sup> T cells expanded for 8 days with IL-2, H9, or H9T. Data are representative of two independent experiments. **b, c**, CD8<sup>+</sup> T cells were isolated and expanded for 8 days with IL-2, H9, or H9T, and 5 million cells were collected for metabolomics analysis; data correspond to this of Fig. 3a. Relative levels of glucose (**b**) and lactate (**c**) are presented as mean values ± SEM, one-way ANOVA test with Dunnett's correction,  $n = 3$  mice. Data are representative of two independent experiments. **d**, 8 day IL-2, H9, or H9T expanded CD8<sup>+</sup> T cells were incubated with or without 2-NBDG to assess glucose uptake. Data are presented as mean values ± SEM, one-way ANOVA test with Dunnett's correction,  $n = 4$  mice. Data are representative of two independent experiments. **e**, Eight-day IL-2, H9, or H9T expanded CD8<sup>+</sup> T cells were incubated with or without TMRM to assess mitochondrial membrane potential. Data are presented as mean values ± SEM, one-way ANOVA test with Dunnett's correction,  $n = 8$  mice. Data are representative of two independent experiments. **f–h**, Pre-activated CD8<sup>+</sup> T cells were treated with control or 1 mM 2-DG in the presence of IL-2 for 2 days. Cells were subsequently stained with antibodies to TCF-1 ( $n = 3$  mice), CD62L ( $n = 3$  mice), or pSTAT5 ( $n = 6$  mice). Data are presented as mean values ± SEM, with two-tailed, paired *t*-test. Data are representative of two independent experiments. **i**, PCA plot of RNA-seq data from 2-DG and control treated cells. Pre-activated CD8<sup>+</sup> T cells were treated with 10 nM IL-2 with or without 2-DG for 2 days, and RNA was extracted for library preparation. Data are from three mice

[Source data](#).

## Extended Data Fig. 5 Comparison of actions of H9T versus natural cytokines in CD8<sup>+</sup> T cells.

**a, b**, Pre-activated mouse (**a**) or human (**b**) CD8<sup>+</sup> T cells were rested and cultured with 10 nM of the indicated cytokines for 6 days, and cell density

was counted by beads-based flow cytometry. Data are mean  $\pm$  SEM, one-way ANOVA test with Dunnett's correction. **a**,  $n = 6$  mice; **b**,  $n=6$  donors. Data are representative of two independent experiments. **c**, Pre-activated mouse CD8 $^{+}$  cells were rested and cultured with 10 nM of the indicated cytokines for 6 days, and TIM-3 expression was examined by flow cytometry. Data are mean  $\pm$  SEM, one-way ANOVA test with Dunnett's correction,  $n = 4$  mice. Data are representative of two independent experiments. **d–f**, Pre-activated mouse CD8 $^{+}$  T cells were rested overnight and cultured with 10 nM of the indicated cytokines for 0, 0.5, 1, 2, or 4 h. Cells were then fixed, permeabilized, and stained for STAT5-pY694 (**d**), AKT-pS473 (**e**) and ERK-pT202/pY204 (**f**). Data are mean  $\pm$  SEM,  $n = 6$  mice, Data are representative of two independent experiments. **g, h**, Pre-activated mouse CD8 $^{+}$  T cells were rested and cultured with 10 nM of the indicated cytokines for 1 day and collected for RNA-seq. PCA analysis (**g**) and selected gene expression (**h**) are shown. Data are combined from two biological repeats

[Source data](#).

[Extended Data Fig. 6 Comparison of H9T with natural cytokines in human CD8 \$^{+}\$  T cells.](#)

**a**, Pre-activated human CD8 $^{+}$  cells were rested and cultured with 10 nM indicated cytokines for 6 days, and TIM-3 expression was examined. Data are mean  $\pm$  SEM, one-way ANOVA test with Dunnett's correction,  $n = 6$  donors. Data are from two independent experiments. **b–d**, Pre-activated human CD8 $^{+}$  T cells were rested and cultured with 10 nM of the indicated cytokines for 0, 0.5, 1, 2 and 4 h and stained with anti-pSTAT5, pAKT, and pERK. Data are mean  $\pm$  SEM,  $n = 6$  donors. Data are from two independent experiments. **e–i**, Pre-activated human CD8 $^{+}$  T cells were rested and cultured with 10 nM of the indicated cytokines for 24 h and cells were collected for RNA-seq. Selected genes expression (**e**), PCA analysis (**f**) and GSEA were shown, Kolmogorov–Smirnov test. Data are from three donors. **j–l**, Pre-activated human CD8 $^{+}$  T cells were rested and cultured with 10 nM of the indicated cytokines for 6 days, and stained with anti-CD27, CCR7 and Granzyme B antibodies. Data are mean  $\pm$  SEM, one-way ANOVA test

with Dunnett's correction,  $n = 6$  donors. Data are from two independent experiments. **m**, Pre-activated human CD8 $^{+}$  cells were rested and cultured with 10 nM of the indicated cytokines for 6 days, followed by TMRM analysis. Data are mean  $\pm$  SEM, one-way ANOVA test with Dunnett's correction,  $n = 5$  donors. Data are from two independent experiments. **n**, Pre-activated human CD8 $^{+}$  T cells were rested and cultured with 1 nM IL-2 alone or in the presence of 1, 10 and 100 nM H9T or IL-15. TIM-3 expression was analysed by flow cytometry after 2 days. Data are mean  $\pm$  SEM, one-way ANOVA test with Dunnett's correction,  $n = 5$  donors. Data are from two independent experiments

[Source data.](#)

### Extended Data Fig. 7 Dose response of IL-2, IL-15 and H9T in CD8 $^{+}$ T cells.

**a–c**, Dose response of IL-2, IL-15, and H9T in human CD8 $^{+}$  T cells. Pre-activated human CD8 $^{+}$  cells were rested and cultured with 0-100 nM of IL-2, IL-15, or H9T for 6 days and stained for surface expression of TIM-3 (**a**) or permeabilized and stained for intracellular granzyme B (**b**). Cell expansion rate (**c**) was assessed using flow cytometry based counting beads. Data are mean  $\pm$  SEM,  $n=2$  donors. Data are representative of two independent experiments. **d–f**, mRNA levels of *TCF7*, *CD27* and *SLC2A1* in human CD8 $^{+}$  T cells. Pre-activated human CD8 $^{+}$  cells were rested and stimulated with 1 nM of IL-2, IL-15, or H9T for 24 h. Cells were collected and mRNA extracted for qPCR analysis of *TCF7* (**d**), *CD27* (**e**) and *SLC2A1* (**f**). The mRNA expression was normalized to that of *RPLP0*. Data are representative of two independent experiments

[Source data.](#)

### Extended Data Fig. 8 Effects of H9T on human CD4 $^{+}$ T cells.

**a–d**, Phenotypic analysis of expanded human CD4 $^{+}$  T cells. Pre-activated human CD4 $^{+}$  cells were cultured with 10 nM H9T, IL-2, IL-15, or IL-7 + IL-15 for 6 days, and mitochondrial membrane potential (**a**) and surface

expression of TIM-3 (**b**), CCR7(**c**), and CD27 (**d**) were examined by flow cytometry analysis after staining. Data are presented as mean values ± SEM,  $n = 5$  donors, with paired two-sided t-test. Data are representative of two independent experiments

[Source data](#).

### **Extended Data Fig. 9 Effects of STAT5 activation on T cell exhaustion and stemness.**

**a**, ChIP-seq analysis of STAT5A- and STAT5B-binding sites at the *Havcr2* locus from previously published datasets (GSE36890). **b, c**, TIM-3 expression on CD8<sup>+</sup> T cells from *Stat5a* (**b**) and *Stat5b* (**c**) knock-out mice versus wild-type littermate controls. Cells were expanded for 8 days with 10 nM of IL-2 as described above, and TIM-3 expression analysed by flow cytometry. Data are mean ± SEM,  $n = 4$  mice, two-sided *t*-test. Data are from two independent repeats. **d**, ChIP-seq analysis of STAT5-binding sites at *HAVCR*, *GZMB*, *TCF7*, *SLC2A1* and *SLC2A3* loci. Human CD8<sup>+</sup> T cells were pre-activated with anti-CD3/anti-CD28 beads for 2 days, rested overnight, incubated with 10 nM IL-2 or H9T for 2 h, and then fixed and lysed for ChIP-seq. Data are from two independent experiments. **e–h**, Three days after retroviral transduction, mouse CD8<sup>+</sup> T cells expressing empty vector (EV)-GFP or STAT5A-1\*6 vector-GFP were sorted and cultured in H9T-containing medium for an additional four days prior to staining with the indicated antibodies. Data are representative of two independent experiments. **i–m**, RNA-seq analysis of the effects of STAT5A-1\*6 expression. Mouse CD8<sup>+</sup> T cells were treated as above, and cells were collected for RNA-seq. PCA analysis (**i**) and expression of selected genes (**j**) are shown. GSEA analysis of IL-2-STAT5 signalling (**k**), PI3K-AKT-mTOR signalling (**l**) and exhaustion versus memory (**m**) are also shown, Kolmogorov-Smirnov test. Data are from two biological repeats. **n–p**, ATAC-seq analysis of the effect of STAT5A-1\*6 expression. Mouse CD8<sup>+</sup> T cells were treated as described above, and cells were collected for ATAC-seq. Shown are PCA analysis to compare IL-2-, H9-, and H9T-expanded cells (**n**), and ATAC-seq data at the *Havcr2* (**o**) and *Tcf7* (**p**) loci

[Source data](#).

## Extended Data Fig. 10 Efficacy of H9T in adoptive cell immunotherapy.

**a**, Tumour growth after transfer of pmel-1 cells that expanded with IL-2, H9 or H9T for 8 days into B16 melanoma-bearing mice, with PBS as a control.  $n = 14$  for H9 group and  $n = 15$  mice for all other groups; data are from three independent repeats. **b**, Blood cells from mice cured of B16 melanoma tumour after adoptive transfer of H9T-expanded CD8<sup>+</sup> pmel-1 cells were stained for CD8 and CD90.1. Data are from two independent experiments. **c, d**, IL-2- or H9T-expanded CD8<sup>+</sup> pmel-1 cells were transferred into B16 melanoma-bearing mice, with PBS as a control. Mice were irradiated one day before cell transfer but not injected i.p. with IL-2 after cell transfer (no IP) or not irradiated but injected with 180,000 IU IL-2 i.p. daily for 3 days beginning on the day of transfer (no IR). Data are mean  $\pm$  SEM,  $n = 5$  mice. Data are from two independent repeats. **e–g**, TIM-3 and PD1 profiling of pmel-1 cells in tumour and draining lymph nodes 7 days after adoptive transfer. Data are mean  $\pm$  SEM,  $n = 5$  mice, one-way ANOVA test with Dunnett's correction. Gating strategy is shown (**g**). Data are from three independent repeats. **h**, B16 tumour size 8 days after pmel-1 CD8<sup>+</sup> T cells infusion. Data are mean  $\pm$  SEM,  $n = 5$  mice. Data are from three independent repeats. **i, j**, Phenotype of pmel-1 cells in tumour and draining lymph nodes 5 or 10 days after adoptive transfer. Data are from two independent experiments. **k–m**, Seven days after adoptive transfer, CD8<sup>+</sup>CD90.1<sup>+</sup> cells was sorted from draining lymph nodes and analysed by RNA-seq. Selected gene expression (**k, l**) and GSEA analysis of memory versus exhausted cells (**m**) with Kolmogorov–Smirnov test are shown. Data are from two independent repeats

[Source data](#).

## Supplementary information

### Supplementary Figures

This file contains Supplementary Figs. 1 and 2. Supplementary Fig. 1 contains the uncropped western blots and Supplementary Fig. 2 shows the

flow cytometry gating strategies.

## **Reporting Summary**

### **Supplementary Table 1**

All RefSeq genes from RNA-seq data related to Fig. 2a.

### **Supplementary Table 2**

Differentially expressed genes between IL-2 and H9T. Related to Fig. 2b, c, and Extended Data Fig. 3a.

### **Supplementary Table 3**

Differentially expressed genes between H9 and H9T. Related to Fig. 2c and Extended Data Fig. 3b, c.

### **Supplementary Table 4**

ATAC-seq peaks of mouse CD8<sup>+</sup> T cells expanded for 6 days as indicated. Related to Fig. 2h–j.

### **Supplementary Table 5**

Significant metabolites. Related to Fig. 3a.

### **Supplementary Table 6**

All detected metabolites. Related to Fig. 3b.

### **Supplementary Table 7**

RNA-seq of mouse CD8<sup>+</sup> T cells treated with IL-2 or IL-2 plus 2-DG for 2 days. Related to Fig. 3m.

## **Supplementary Table 8**

RNA-seq of mouse CD8<sup>+</sup> T cells treated with H9T, IL-2, IL-15 or IL-7 + 15 for 24 hours. Related to Extended Data Fig. 5h.

## **Supplementary Table 9**

RNA-seq of human CD8<sup>+</sup> T cells treated with H9T, IL-2, IL-15 or IL-7 + 15 for 24 hours. Related to Extended Data Fig. 6e.

## **Supplementary Table 10**

RNA-seq of mouse CD8<sup>+</sup> T cells overexpressing STAT5A-1\*6 or a control vector. Related to Extended Data Fig. 9j.

## **Supplementary Table 11**

ATAC-seq of mouse CD8<sup>+</sup> T cells overexpressing STAT5A-1\*6 or a control vector. Related to Extended Data Fig. 9n.

## **Supplementary Table 12**

RNA-seq of CD90.1<sup>+</sup> mouse CD8<sup>+</sup> T cells purified from draining lymph nodes 7 days after transfer. Related to Extended Data Fig. 10k.

## **Source data**

### **Source Data Fig. 1**

### **Source Data Fig. 2**

### **Source Data Fig. 3**

### **Source Data Fig. 4**

[Extended Data Source Data Fig. 1](#)

[Extended Data Source Data Fig. 2](#)

[Extended Data Source Data Fig. 3](#)

[Extended Data Source Data Fig. 4](#)

[Extended Data Source Data Fig. 5](#)

[Extended Data Source Data Fig. 6](#)

[Extended Data Source Data Fig. 7](#)

[Extended Data Source Data Fig. 8](#)

[Extended Data Source Data Fig. 9](#)

[Extended Data Source Data Fig. 10](#)

## Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Mo, F., Yu, Z., Li, P. *et al.* An engineered IL-2 partial agonist promotes CD8<sup>+</sup> T cell stemness. *Nature* **597**, 544–548 (2021).  
<https://doi.org/10.1038/s41586-021-03861-0>

- Received: 05 October 2019
- Accepted: 23 July 2021

- Published: 15 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03861-0>

## 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/s41586-021-03861-0>

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

- Article
- [Published: 08 September 2021](#)

# Inter-cellular CRISPR screens reveal regulators of cancer cell phagocytosis

- [Roarke A. Kamber](#)<sup>1</sup>,
- [Yoko Nishiga](#)<sup>1,2,3</sup>,
- [Bhek Morton](#)<sup>1</sup>,
- [Allison M. Banuelos](#)<sup>4,5,6</sup>,
- [Amira A. Barkal](#)<sup>4,5,6,7</sup>,
- [Felipe Vences-Catalán](#)<sup>8</sup>,
- [Mingxin Gu](#)<sup>1</sup>,
- [Daniel Fernandez](#) ORCID: [orcid.org/0000-0002-6221-152X](https://orcid.org/0000-0002-6221-152X)<sup>9,10</sup>,
- [Jose A. Seoane](#) ORCID: [orcid.org/0000-0002-3856-9177](https://orcid.org/0000-0002-3856-9177)<sup>6,11</sup>,
- [David Yao](#)<sup>1</sup>,
- [Katherine Liu](#)<sup>1</sup>,
- [Sijie Lin](#)<sup>1</sup>,
- [Kaitlyn Spees](#)<sup>1</sup>,
- [Christina Curtis](#)<sup>1,6,11,12</sup>,
- [Livnat Jerby-Arnon](#) ORCID: [orcid.org/0000-0002-4037-386X](https://orcid.org/0000-0002-4037-386X)<sup>1,13</sup>,
- [Irving L. Weissman](#) ORCID: [orcid.org/0000-0002-9077-7467](https://orcid.org/0000-0002-9077-7467)<sup>4,5,6,14</sup>,
- [Julien Sage](#) ORCID: [orcid.org/0000-0002-8928-9968](https://orcid.org/0000-0002-8928-9968)<sup>1,2,6,9</sup> &
- [Michael C. Bassik](#) ORCID: [orcid.org/0000-0001-5185-8427](https://orcid.org/0000-0001-5185-8427)<sup>1,6,9</sup>

*Nature* volume **597**, pages 549–554 (2021)

- 11k Accesses

- 112 Altmetric
- [Metrics details](#)

## Subjects

- [Cancer](#)
- [Genetics](#)

## Abstract

Monoclonal antibody therapies targeting tumour antigens drive cancer cell elimination in large part by triggering macrophage phagocytosis of cancer cells<sup>1,2,3,4,5,6,7</sup>. However, cancer cells evade phagocytosis using mechanisms that are incompletely understood. Here we develop a platform for unbiased identification of factors that impede antibody-dependent cellular phagocytosis (ADCP) using complementary genome-wide CRISPR knockout and overexpression screens in both cancer cells and macrophages. In cancer cells, beyond known factors such as CD47, we identify many regulators of susceptibility to ADCP, including the poorly characterized enzyme adipocyte plasma membrane-associated protein (APMAP). We find that loss of APMAP synergizes with tumour antigen-targeting monoclonal antibodies and/or CD47-blocking monoclonal antibodies to drive markedly increased phagocytosis across a wide range of cancer cell types, including those that are otherwise resistant to ADCP. Additionally, we show that APMAP loss synergizes with several different tumour-targeting monoclonal antibodies to inhibit tumour growth in mice. Using genome-wide counterscreens in macrophages, we find that the G-protein-coupled receptor GPR84 mediates enhanced phagocytosis of APMAP-deficient cancer cells. This work reveals a cancer-intrinsic regulator of susceptibility to antibody-driven phagocytosis and, more broadly, expands our knowledge of the mechanisms governing cancer resistance to macrophage phagocytosis.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

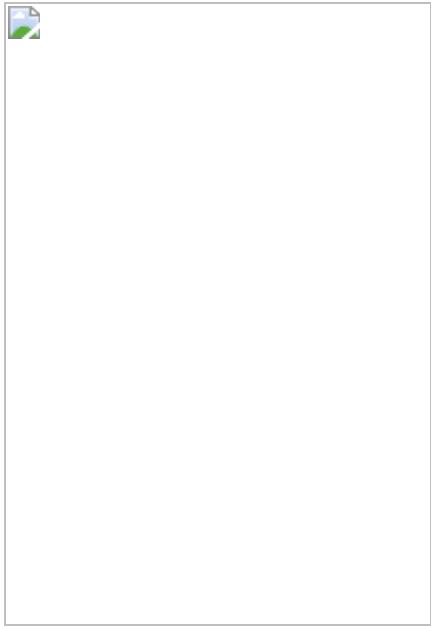
[Rent or Buy](#)

All prices are NET prices.

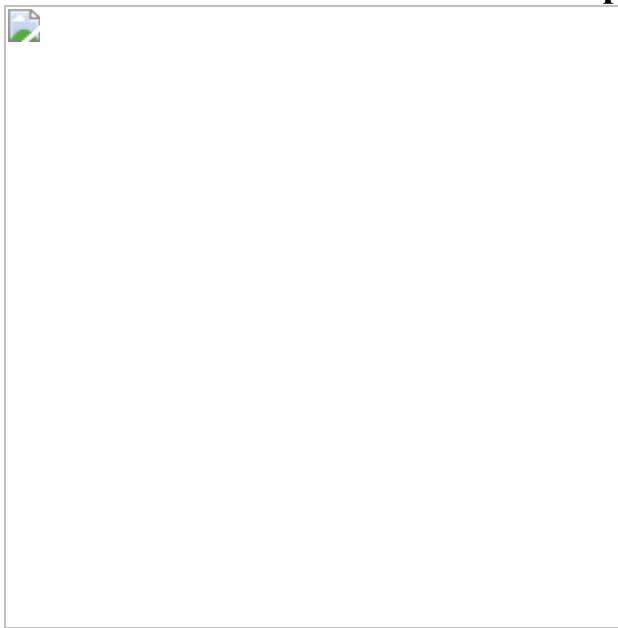
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

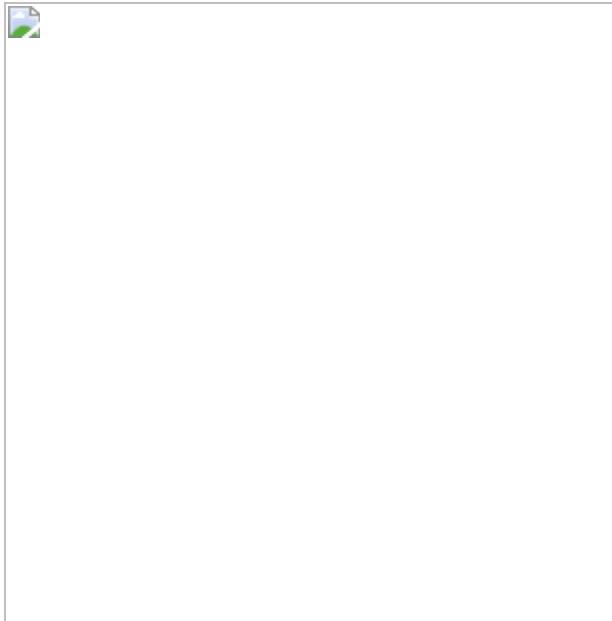
**Fig. 1: Genome-wide CRISPR screens reveal novel regulators of ADCP.**



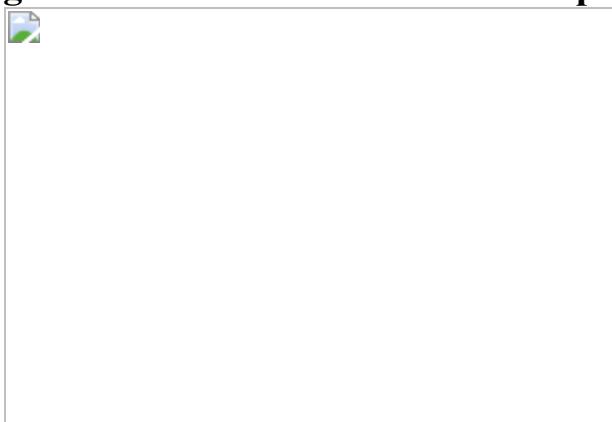
**Fig. 2: APMAP loss synergizes with monoclonal antibodies and CD47 blockade to increase cancer cell susceptibility to phagocytosis.**



**Fig. 3: APMAP loss sensitizes diverse tumour types to monoclonal antibodies in vitro and in mice.**



**Fig. 4: GPR84 mediates enhanced uptake of APMAP<sup>KO</sup> cancer cells.**



## Data availability

CRISPR screen and RNA-seq raw sequencing data are available under BioProject accession number [PRJNA748551](#). All other primary data for all figures and supplementary figures are available from the corresponding author upon request. Gene dependency data from the Cancer Dependency Map are publicly available at [www.depmap.org](http://www.depmap.org). Cancer expression data from The Cancer Genome Atlas are available at <https://gdc.cancer.gov>. CCLE data are available at <https://sites.broadinstitute.org/ccle/>. Source data are provided with this paper.

# References

1. 1.  
Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **12**, 278–287 (2012).
2. 2.  
Sliwkowski, M. X. & Mellman, I. Antibody therapeutics in cancer. *Science* **341**, 1192–1198 (2013).
3. 3.  
Weiskopf, K. & Weissman, I. L. Macrophages are critical effectors of antibody therapies for cancer. *mAbs* **7**, 303–310 (2015).
4. 4.  
Tsao, L.-C. et al. CD47 blockade augmentation of trastuzumab antitumor efficacy dependent on antibody-dependent cellular phagocytosis. *JCI Insight* **4**, e131882 (2019).
5. 5.  
Brodsky, F. M. Monoclonal antibodies as magic bullets. *Pharm. Res.* **5**, 1–9 (1988).
6. 6.  
Maleki, L. A., Baradaran, B., Majidi, J., Mohammadian, M. & Shahneh, F. Z. Future prospects of monoclonal antibodies as magic bullets in immunotherapy. *Hum. Antibodies* **22**, 9–13 (2013).
7. 7.  
Glennie, M. J., French, R. R., Cragg, M. S. & Taylor, R. P. Mechanisms of killing by anti-CD20 monoclonal antibodies. *Mol.*

*Immunol.* **44**, 3823–3837 (2007).

8. 8.

Chao, M. P. et al. Anti-CD47 antibody synergizes with rituximab to promote phagocytosis and eradicate non-Hodgkin lymphoma. *Cell* **142**, 699–713 (2010).

9. 9.

Logtenberg, M. E. W. et al. Glutaminyl cyclase is an enzymatic modifier of the CD47–SIRP $\alpha$  axis and a target for cancer immunotherapy. *Nat. Med.* **25**, 612–619 (2019).

10. 10.

Macaulay, M. S., Crocker, P. R. & Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nat. Rev. Immunol.* **14**, 653–666 (2014).

11. 11.

Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

12. 12.

Gao, S. et al. The oncogenic role of MUC12 in RCC progression depends on c-Jun/TGF- $\beta$  signalling. *J. Cell. Mol. Med.* **24**, 8789–8802 (2020).

13. 13.

Taylor-Papadimitriou, J. et al. MUC1 and the immunobiology of cancer. *J. Mammary Gland Biol. Neoplasia* **7**, 209–221 (2002).

14. 14.

O'Prey, J., Wilkinson, S. & Ryan, K. M. Tumor antigen LRRC15 impedes adenoviral infection: implications for virus-based cancer therapy. *J. Virol.* **82**, 5933–5939 (2008).

15. 15.

Purcell, J. W. et al. LRRC15 is a novel mesenchymal protein and stromal target for antibody–drug conjugates. *Cancer Res.* **78**, 4059–4072 (2018).

16. 16.

Itoh, Y. et al. Identification and expression of human epiglycanin/MUC21: a novel transmembrane mucin. *Glycobiology* **18**, 74–83 (2008).

17. 17.

Snyder, K. A. et al. Podocalyxin enhances breast tumor growth and metastasis and is a target for monoclonal antibody therapy. *Breast Cancer Res.* **17**, 46 (2015).

18. 18.

Tarbé, N. G., Rio, M.-C., Hummel, S., Weidle, U. H. & Zöller, M. Overexpression of the small transmembrane and glycosylated protein SMAGP supports metastasis formation of a rat pancreatic adenocarcinoma line. *Int. J. Cancer* **117**, 913–922 (2005).

19. 19.

Ajona, D. et al. Blockade of the complement C5a/C5aR1 axis impairs lung cancer bone metastasis by CXCL16-mediated effects. *Am. J. Respir. Crit. Care Med.* **197**, 1164–1176 (2018).

20. 20.

Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60 (2004).

21. 21.

Jiang, S. et al. Cholesterol induces epithelial-to-mesenchymal transition of prostate cancer cells by suppressing degradation of EGFR through APMAP. *Cancer Res.* **79**, 3063–3075 (2019).

22. 22.

Gerber, H. et al. The APMAP interactome reveals new modulators of APP processing and beta-amyloid production that are altered in Alzheimer's disease. *Acta Neuropathol. Commun.* **7**, 13 (2019).

23. 23.

Ilhan, A. et al. Localization and characterization of the novel protein encoded by C20orf3. *Biochem. J.* **414**, 485–495 (2008).

24. 24.

Barkal, A. A. et al. Engagement of MHC class I by the inhibitory receptor LILRB1 suppresses macrophages and is a target of cancer immunotherapy. *Nat. Immunol.* **19**, 76–84 (2018).

25. 25.

Clynes, R., Takechi, Y., Moroi, Y., Houghton, A. & Ravetch, J. V. Fc receptors are required in passive and active immunity to melanoma. *Proc. Natl Acad. Sci. USA* **95**, 652–656 (1998).

26. 26.

Lattin, J. E. et al. Expression analysis of G protein-coupled receptors in mouse macrophages. *Immunome Res.* **4**, 5 (2008).

27. 27.

Recio, C. et al. Activation of the immune-metabolic receptor GPR84 enhances inflammation and phagocytosis in macrophages. *Front. Immunol.* **9**, 1419 (2018).

28. 28.

Wang, J., Wu, X., Simonavicius, N., Tian, H. & Ling, L. Medium-chain fatty acids as ligands for orphan G protein-coupled receptor GPR84. *J. Biol. Chem.* **281**, 34457–34464 (2006).

29. 29.

Gont, A., Daneshmand, M., Woulfe, J. & Lorimer, I. PREX1 integrates G protein-coupled receptor and phosphoinositide 3-kinase signaling to promote glioblastoma invasion. *Eur. J. Cancer* **61**, S171–S172 (2016).

30. 30.

Noy, R. & Pollard, J. W. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* **41**, 49–61 (2014).

31. 31.

Cunha, L. D. et al. LC3-associated phagocytosis in myeloid Cells Promotes Tumor Immune Tolerance. *Cell* **175**, 429–441.e16 (2018).

32. 32.

Su, S. et al. Immune checkpoint inhibition overcomes ADCP-induced immunosuppression by macrophages. *Cell* **175**, 442–457.e23 (2018).

33. 33.

Pathria, P., Louis, T. L. & Varner, J. A. Targeting tumor-associated macrophages in cancer. *Trends Immunol.* **40**, 310–327 (2019).

34. 34.

Ruffell, B. & Coussens, L. M. Macrophages and therapeutic resistance in cancer. *Cancer Cell* **27**, 462–472 (2015).

35. 35.

Hicks, M. A. et al. The evolution of function in strictosidine synthase-like proteins. *Proteins Struct. Funct. Bioinf.* **79**, 3082–3098 (2011).

36. 36.

Khersonsky, O. & Tawfik, D. S. Structure-reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase. *Biochemistry* **44**, 6371–6382 (2005).

37. 37.

Flannagan, R. S., Jaumouillé, V. & Grinstein, S. The cell biology of phagocytosis. *Annu. Rev. Pathol.* **7**, 61–98 (2012).

38. 38.

Manguso, R. T. et al. In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413–418 (2017).

39. 39.

Lawson, K. A. et al. Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature* **586**, 120–126 (2020).

40. 40.

Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).

41. 41.

Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).

42. 42.

Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).

43. 43.

Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).

44. 44.

Jeng, E. E. et al. Systematic identification of host cell regulators of *Legionella pneumophila* pathogenesis using a genome-wide CRISPR screen. *Cell Host Microbe* **26**, 551–563.e6 (2019).

45. 45.

Haney, M. S. et al. Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat. Genet.* **50**, 1716–1727 (2018).

46. 46.

Reimand, J. et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).

47. 47.

Schutze, M.-P., Peterson, P. A. & Jackson, M. R. An N-terminal double-arginine motif maintains type II membrane proteins in the endoplasmic reticulum. *EMBO J.* **13**, 1696–1705 (1994).

48. 48.

Delaveris, C. S., Chiu, S. H., Riley, N. M. & Bertozzi, C. R. Modulation of immune cell reactivity with *cis*-binding Siglec agonists. *Proc. Natl Acad. Sci. USA* **118**, e2012408118 (2021).

49. 49.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

50. 50.

Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).

51. 51.

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).

52. 52.

Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

53. 53.

Jerby-Arnon, L. et al. Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).

54. 54.

Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997.e24 (2018).

55. 55.

Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013.e20 (2018).

56. 56.

Neftel, C. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e21 (2019).

57. 57.

Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

58. 58.

Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence–structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243–257 (2001).

59. 59.

Ben-David, M. et al. Catalytic versatility and backups in enzyme active sites: the case of serum paraoxonase 1. *J. Mol. Biol.* **418**, 181–196 (2012).

60. 60.

Tanaka, Y. et al. Structural and mutational analyses of Drp35 from *Staphylococcus aureus*: a possible mechanism for its lactonase activity. *J. Biol. Chem.* **282**, 5770–5780 (2007).

61. 61.

Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

62. 62.

Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).

63. 63.

Sockolosky, J. T. et al. Durable antitumor responses to CD47 blockade require adaptive immune stimulation. *Proc. Natl Acad. Sci. USA* **113**, E2646–E2654 (2016).

## Acknowledgements

We thank R. Levy, S. Levy, C. Bertozzi, J. Long, S. Dixon, P. Jackson, M. Smith, T. Wyss-Coray, A. Drainas, A. Derry, B. Smith, C. Delaveris, J. Shon, S. Wisnovsky, J. Donnelly, E. Zhang, T. Raveh, D. Vorselen, J. Carozza, M. Ellenberger, J. Chan, L. Jiang, R. Jian, M. Snyder, K. McNamara, R. Chen and members of the Bassik laboratory, including G. Hess, R. Levin, K. Han, K. Tsui, M. Haney, D. Morgens, J. Tycko, M. Dubreuil, K. Aloul, B. Ego and A. Li, for helpful discussions and experimental advice; and A. Sil for providing the J774 Cas9 line. Cell sorting for this project was done on instruments in the Stanford Shared FACS Facility, including an instrument purchased by the Parker Institute for Cancer Immunotherapy. This research was supported by an NIH Director’s New Innovator award (1DP2HD084069-01) to M.C.B., by the Ludwig Institute for Cancer Research (J.S. and I.W.), the NIH (grants CA213273 and CA231997 to J.S., R35CA220434-05 and 1R01AI143889-01A1 to I.W.), the JSPS (JSPS overseas research fellowship to Y.N.), a Stanford School of Medicine Dean’s Postdoctoral Fellowship to R.A.K. and a Jane Coffin Childs Postdoctoral Fellowship to R.A.K. L.J.A. is a Chan Zuckerberg Biohub Investigator and holds a Career Award at the Scientific Interface from BWF.

## Author information

### Affiliations

1. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

Roarke A. Kamber, Yoko Nishiga, Bhek Morton, Mingxin Gu, David Yao, Katherine Liu, Sijie Lin, Kaitlyn Spees, Christina Curtis, Livnat Jerby-Arnon, Julien Sage & Michael C. Bassik

2. Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

Yoko Nishiga & Julien Sage

3. Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA

Yoko Nishiga

4. Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

Allison M. Banuelos, Amira A. Barkal & Irving L. Weissman

5. Ludwig Center for Cancer Stem Cell Research and Medicine, Stanford University School of Medicine, Stanford, CA, USA

Allison M. Banuelos, Amira A. Barkal & Irving L. Weissman

6. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

Allison M. Banuelos, Amira A. Barkal, Jose A. Seoane, Christina Curtis, Irving L. Weissman, Julien Sage & Michael C. Bassik

7. Stanford Medical Scientist Training Program, Stanford University, Stanford, CA, USA

Amira A. Barkal

8. Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

Felipe Vences-Catalán

9. Program in Chemistry, Engineering, and Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA, USA

Daniel Fernandez, Julien Sage & Michael C. Bassik

10. Stanford ChEM-H, Macromolecular Structure Knowledge Center,  
Stanford University, Stanford, CA, USA

Daniel Fernandez

11. Department of Medicine, Stanford University School of Medicine,  
Stanford, CA, USA

Jose A. Seoane & Christina Curtis

12. Program in Cancer Biology, Stanford University School of Medicine,  
Stanford, CA, USA

Christina Curtis

13. Chan Zuckerberg Biohub, San Francisco, CA, USA

Livnat Jerby-Arnon

14. Department of Pathology, Stanford University School of Medicine,  
Stanford, CA, USA

Irving L. Weissman

## Contributions

R.A.K. and M.C.B. conceived and designed the study. R.A.K. designed the cancer–macrophage co-culture system for genome-wide CRISPR screens. R.A.K. performed the CRISPR screens in Ramos cells and J774 cells with help from S.L. and K.S., and B.M. performed the CRISPR screens in Karpas-299 cells. Y.N. performed *in vivo* mouse experiments in NSG mice, with advice from J.S. A.M.B. and A.A.B. performed the syngeneic mouse experiments with advice from I.L.W. and F.V.-C. D.F. generated the APMAP homology model. J.A.S. analysed the TCGA data for differential expression in different cancer types, with advice from C.C. L.J.-A. analysed single-cell RNA-sequencing data. R.A.K. and M.G. performed Incucyte

assays to validate CRISPR knockout hits. R.A.K., M.G. and S.L. cloned sgRNA vectors and generated knockout cell lines. R.A.K. performed the western blots, confocal microscopy and drug titrations. M.G., S.L. and R.A.K. performed flow cytometry analyses. R.A.K. and S.L. performed RNA-sequencing, and D.Y. and K.L. analysed the RNA-sequencing data. D.Y. helped with design of the oligonucleotide sub-libraries and K.S. cloned the sub-libraries. R.A.K. and M.C.B. wrote the manuscript. All authors discussed the results and the manuscript.

## Corresponding author

Correspondence to Michael C. Bassik.

## Ethics declarations

## Competing interests

R.A.K. and M.C.B., through the Office of Technology Licensing at Stanford University, have filed a patent application on the methods and findings in this manuscript. I.W. is an inventor on several patents in the field of ADCP induced by blockade of several don't-eat-me signals such as CD47, CD24, beta-2-microglobulin, and PDL1, and their macrophage cognate receptors, respectively, SIRPa, Siglec-10, LILRB1, and PD1. These have been licensed to several companies. I.W. is not currently affiliated with these companies and does not hold stock in them. He is, however, engaged in the formation of one or more start-up companies in the field. J.S. licensed a patent to Forty Seven Inc./Gilead on the use of CD47 blocking strategies in SCLC.

## Additional information

**Peer review information** *Nature* thanks Ross Levine and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 CRISPR knockout screening platform for regulators of ADCP in cancer cells and batch retest validation screen results.](#)

**a**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells by J774 macrophages in the presence or absence of anti-CD20 and/or anti-CD47 antibodies. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in untreated control condition at final timepoint. Data represent mean  $\pm$  s.d. ( $n = 4$ ). Two-way ANOVA with Bonferroni correction. **b**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells by J774 or U937 macrophages in the presence of anti-CD20. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in U937 cells at final timepoint. Data represent mean  $\pm$  s.d. ( $n = 3$ ). Two-way ANOVA with Bonferroni correction. **c**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells by J774 macrophages with or without 24 h pre-treatment with 100 ng ml<sup>-1</sup> LPS. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in untreated control condition at final timepoint. Data represent mean  $\pm$  s.d. ( $n = 4$ ). Two-way ANOVA with Bonferroni correction. **d**, Differential expression analysis of J774 macrophages before and after treatment (24 h) with 100 ng ml<sup>-1</sup> LPS, showing induction of classic LPS-activated M1 macrophage markers NOS2 and IL1B. **e**, Replicates for CRISPRko screen in Ramos cells for susceptibility to ADCP driven by anti-CD20 antibodies. **f**, Gene ontology enrichment analysis for negative Ramos CRISPRko ADCP hits (cutoff of CasTLE score  $> 50$ ). n indicates number of genes among query gene list annotated with indicated term. **g**, Batch re-test screen for ADCP sensitivity in Ramos Cas9 cells. Library comprised top 250 hits (both positive and negative effect sizes) from genome-wide CRISPRko screen and top 480 anti-phagocytic hits from CRISPRa screen. Hits were defined based on 95% confidence interval of CasTLE effect size (see

Supplementary Table 3). **h**, Replicates of Batch re-test screen in Ramos Cas9 cells. **i**, Comparison of Ramos batch re-test and genome-wide ADCP CRISPRko screens for genes that were hits in the CRISPRko screen. **j**, Survival assay for Ramos Cas9 cells subjected to treatment with macrophages and anti-CD20, expressing indicated sgRNAs (2 distinct sgRNAs per gene). GFP<sup>+</sup> Ramos Cas9 cells expressing negative control sgRNA were mixed with an equal number of mCherry<sup>+</sup> cells expressing indicated sgRNAs and cultured in the presence of J774 macrophages and anti-CD20 antibodies. Plotted is the mean percentage of surviving Ramos cells that were mCherry<sup>+</sup> after 2 d, normalized to control (Ctrl) Ramos cells that expressed an empty vector) ( $n = 3$  cell culture wells, data represent mean  $\pm$  s.d.). One-way ANOVA with Bonferroni correction.

### Extended Data Fig. 2 CRISPR activation screening platform development and analysis of anti-phagocytic hits.

**a**, Validation of Ramos CRISPRa clones. Single-cell derived CRISPRa clones were constructed as described in the Methods and transduced with sgRNAs targeting CD2 (using a lentiviral vector co-expressing GFP). Indicated clones and parent Ramos cells were stained with anti-CD2-APC antibodies. Mean APC signal in the GFP<sup>+</sup> population is plotted ( $n = 2$  technical replicates, mean is shown). Clone #6 was used for screening. **b**, Replicates for CRISPRa screen in Ramos cells for susceptibility to ADCP driven by anti-CD20 and anti-CD47 antibodies. **c**, Gene ontology enrichment analysis for positive Ramos CRISPRa ADCP hits (cutoff of CasTLE score  $> 50$ ) (top) and top 50 anti-phagocytic factors (bottom).  $n$  indicates number of genes among query gene list annotated with that term. **d**, Schematic of time-lapse imaging assay for ADCP. pHrodo-Red fluorescence intensity increases in low-pH conditions, such as in the lysosome following internalization of the target cell. **e**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells, stably expressing indicated constructs, by J774 macrophages in the presence of anti-CD20 and anti-CD47 antibodies. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in GFP-FLAG cells at final timepoint. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). Two-way ANOVA with Bonferroni correction. **f**, Expression (TPM) of

*SMAGP* in 1304 cell lines in CCLE. **g, h**, Flow cytometry assays for anti-CD20 and anti-CD45 binding to Ramos CRISPRa cells expressing indicated sgRNAs. Data represent mean  $\pm$  s.d. ( $n = 3$  independently stained samples). One-way ANOVA with Bonferroni correction. **i**, Volcano plot of screen in Karpas-299 cells conducted in presence of anti-CD30 antibodies. Dotted line indicates 5% FDR. **j**, Heatmap of differential expression for 12 selected anti-phagocytic genes in 23 tumour types compared to normal tissue. Tumour type abbreviations are listed here: <https://gdc.cancer.gov/resources-tcga-users/tcg-a-code-tables/tcg-a-study-abbreviations>. **k**, Minimum expression (TPM) across all cell lines in CCLE is plotted against maximum probability of essentiality in all cell lines profiled in DepMap for 50 top anti-phagocytic hits shown in Fig. 1e.

### **Extended Data Fig. 3 Screens for cancer cell regulators of ADCP in the presence or absence of CD47 and evaluation of importance of antibodies and Fc receptor for APMAP effect.**

**a**, Schematic and volcano plot of CRISPR screen in Ramos Cas9 cells for sensitivity to macrophage phagocytosis in the presence of anti-CD20 in cells expressing an sgRNA targeting a Safe locus. Dotted line indicates 5% FDR. A transmembrane gene-enriched sublibrary containing 3,124 genes was used. **b**, Schematic and volcano plot of CRISPR screen in Ramos Cas9 cells for sensitivity to macrophage phagocytosis in the presence of anti-CD20 in cells expressing an sgRNA targeting the *CD47* locus. Dotted line indicates 5% FDR. A transmembrane gene-enriched sublibrary containing 3,124 genes was used. **c**, Schematic and volcano plot of CRISPRko screen in Ramos Cas9 cells for sensitivity to macrophage phagocytosis in the presence of anti-CD20 and anti-CD47 in cells expressing an sgRNA targeting a Safe locus. Dotted line indicates 5% FDR. A transmembrane gene-enriched sublibrary containing 3,124 genes was used. **d**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells with indicated genotypes by human primary peripheral blood-derived macrophages, from two independent healthy de-identified human donors, in the presence or absence of anti-CD47 antibodies. Phagocytosis index normalized to control (Safe<sup>KO</sup>) cells without anti-CD47. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). One-way ANOVA with Bonferroni correction. **e**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells

with indicated genotypes by J774 macrophages in the presence of anti-CD20 or anti-CD47 antibodies. Where indicated, J774 macrophages were pre-incubated with Fc-blocking antibodies for 45 min on ice. Phagocytosis index normalized to control ( $\text{Safe}^{\text{KO}}$ ) cells without antibody analysed in parallel (condition not shown). Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). Two-way ANOVA with Bonferroni correction. **f**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells with indicated genotypes by J774 macrophages in the absence of antibodies. Phagocytosis index normalized to control ( $\text{Safe}^{\text{KO}}/\text{Safe}^{\text{KO}}$ ) cells. Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). One-way ANOVA with Bonferroni correction.

**Extended Data Fig. 4 APMAP loss sensitizes cells to ADCP in a highly specific manner and without affecting surface levels of other pro- and anti-phagocytic factors.**

**a**, Phagocytosis assay for uptake of pHrodo-labelled Karpas-299 Cas9 cells expressing indicated sgRNAs by J774 macrophages in the presence or absence of anti-CD30 antibodies. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in untreated control condition at final timepoint. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). Two-way ANOVA with Bonferroni correction. **b**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells with indicated genotypes by human U937 macrophages in the presence or absence of anti-CD20 (rituximab) antibodies at indicated concentrations. Phagocytosis index normalized to control ( $\text{Safe}^{\text{KO}}$ ) Ramos cells without anti-CD20. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). Two-way ANOVA with Bonferroni correction. **c**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells with indicated genotypes by human primary peripheral blood-derived macrophages, from two independent healthy de-identified human donors, in the presence or absence of  $10 \text{ ng ml}^{-1}$  anti-CD20 antibodies. Phagocytosis index normalized to control ( $\text{Safe}^{\text{KO}}$ ) Ramos cells without anti-CD20. Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). One-way ANOVA with Bonferroni correction. **d**, Phagocytosis assay for uptake of pHrodo-labelled Ramos Cas9 cells expressing indicated sgRNAs by J774 macrophages with or without 24 h

pre-treatment with 100 ng ml<sup>-1</sup> LPS. Normalized phagocytosis index was calculated as average total pHrodo Red signal per well, normalized to signal in untreated control condition at final timepoint. Data represent mean  $\pm$  s.d. (n = 4 cell culture wells). **e**, Ramos Cas9 cells expressing indicated sgRNAs were incubated with Annexin V-FITC or anti-Calreticulin-DyLight-488 and analysed by flow cytometry. CRT, calreticulin. Data represent mean  $\pm$  s.d. (n = 3 independently stained samples). P-values were from two-tailed t-tests. **f**, Flow-cytometry based measurement of cell surface levels of CD20 in Ramos Cas9 cells expressing indicated sgRNAs. Data represent mean (n = 2 independently stained samples, except cells expressing CD20 sgRNA (n = 1)). **g**, Flow-cytometry based measurement of cell surface levels of CD47 in Ramos Cas9 cells expressing indicated sgRNAs. Data represent mean  $\pm$  s.d. (n = 3 independently stained samples). **h**, Flow-cytometry based measurement of cell surface levels of sialic acid in Ramos Cas9 cells expressing indicated sgRNAs. Where indicated, cells were treated with sialidase as a positive control. Data represent mean  $\pm$  s.d. (n = 3 independently stained samples). **i**, Viability assays (measured as cell confluence after 72 h on Incucyte, normalized to untreated Safe<sup>KO</sup> control cells) of indicated Ramos cells in the presence of indicated concentrations of 9 drugs. Data represent mean  $\pm$  s.d. (n = 3 cell culture wells). **j**, Flow-cytometry based measurement of forward scatter (FSC) and side scatter (SSC) in Ramos Cas9 cells expressing indicated sgRNAs. Data represent mean  $\pm$  s.d. (n = 3 independently analysed samples). **k**, Ramos-J774 adhesion assay in the presence of indicated antibody concentrations, using indicated GFP+ Ramos Cas9 knockout cells. Data represent mean  $\pm$  s.d. (n = 2 cell culture wells). **l**, Flow-cytometry based measurement of ADCP of Ramos Cas9 cells expressing indicated sgRNAs and stained with either calcein or CellTrace-Far-Red dye before incubation with J774 macrophages and anti-CD20 for 24h. Data represent mean  $\pm$  s.d. (n = 3 cell culture wells). Two-tailed t-tests were used to compare Safe<sup>KO</sup> and APMAP<sup>KO</sup> cells within each labeling condition.

**Extended Data Fig. 5 APMAP localizes to the endoplasmic reticulum and its cytosolic domain, transmembrane domain, and N-glycosylation are not required for its function in ADCP.**

**a**, Localization of APMAP-FLAG and APMAP<sup>E103A</sup>-FLAG to the endoplasmic reticulum in HeLa cells. Scale bar, 20 μm. Calnexin is used as a marker of the endoplasmic reticulum. FLAG staining was representative of two independent experiments. **b**, Immunoblotting of cell extracts derived from Ramos cells of indicated genotypes expressing indicated APMAP-FLAG constructs. GAPDH served as loading control. Experiment was performed twice. **c, d**, Phagocytosis assay for uptake of pHrodo-labelled Ramos-Cas9 cells with indicated genotypes by J774 macrophages in the presence of anti-CD20 antibodies. APMAP-F, APMAP-FLAG. TFRC<sup>RR</sup>, mutant allele of TFRC that localizes primarily to the endoplasmic reticulum<sup>47</sup>. Phagocytosis index normalized to control (Safe<sup>KO</sup>) cells. Data represent mean +/- s.d. (n = 4 cell culture wells). One-way ANOVA with Bonferroni correction. **e**, Immunoblotting of cell extracts that were treated, where indicated, with PNGase F to remove N-glycosylation. Actin served as loading control. Experiment was performed once. **f**, Phagocytosis assay for uptake of pHrodo-labelled Ramos Cas9 cells expressing indicated sgRNAs and indicated addback constructs by J774 macrophages in the presence of anti-CD20 antibodies. Normalized phagocytosis index was calculated as average total pHrodo Red signal at 5 h for each well, normalized to signal in Safe<sup>KO</sup> cells at 5 h timepoint. Data represent mean +/- s.d. (n = 3 cell culture wells). One-way ANOVA with Bonferroni correction. For gel source data, see Supplementary Figure 1.

### **Extended Data Fig. 6 Evaluation of APMAP role in ADCP across diverse cancer cell lines and in syngeneic mice.**

**a**, Levels of APMAP in ten cell lines measured by Western blot. All cell lines stably express Cas9 and were transduced with indicated sgRNAs. Actin served as loading control. Western blot to confirm knockout across all ten cell lines on one gel was performed once. For gel source data, see Supplementary Figure 1. **b**, Expression levels (TPM) of CD47 and APMAP in ten cell lines (data from CCLE). **c**, Survival measurements of selected (GFP<sup>+</sup>) cell lines in Fig. 3a, measured as percentage of GFP remaining after indicated number of hours of incubation with J774 macrophages in presence or absence of anti-CD47. Data represent mean +/- s.d. (n = 4 cell culture wells, except Karpas-299 (n = 3)). One-way ANOVA with multiple

comparisons correction. **d**, Phagocytosis assays as in Fig. [3a](#), but with isotype control antibodies. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). One-way ANOVA with Bonferroni correction. **e**, Phagocytosis assay for uptake of pHrodo-labelled cells for indicated Cas9-expressing cell lines expressing indicated sgRNAs by J774 macrophages in the presence or absence of anti-EGFR/cetuximab antibodies. Phagocytosis index normalized to control (Safe<sup>KO</sup>) cells without anti-EGFR. Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). One-way ANOVA with Bonferroni correction. **f**, Survival measurements of selected (GFP<sup>+</sup>) cell lines in Extended Data Fig. [6e](#), measured as percentage of GFP remaining after indicated number of hours of incubation with J774 macrophages in presence or absence of anti-EGFR. Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). One-way ANOVA with Bonferroni correction. **g**, Representative photographs depicting Ramos tumours of indicated genotype extracted from NSG mice at 25 d following transplantation. **h**, Safe<sup>KO</sup> or APMAP<sup>KO</sup> Ramos cells were transplanted into NSG mice and allowed to form tumours. Mice were treated with anti-CD47 (B6H12, BioXCell) or PBS daily starting 17 d following transplantation, and tumour size was measured every 2 d. Data represent mean  $\pm$  s.e.m. ( $n = 5$  (Safe<sup>KO</sup> groups) and 6 (APMAP<sup>KO</sup> groups)). Two-way ANOVA with Tukey correction (comparison between Safe<sup>KO</sup>/anti-CD47 and APMAP<sup>KO</sup>/anti-CD47 for final timepoint is shown). **i**, Mouse weights in Ramos (top) and NCI-H82 (bottom) xenograft experiments (Extended Data Fig. [6h](#), Fig. [3b](#)). Data represent mean  $\pm$  s.d. Two-way ANOVA with Bonferroni correction ( $n = 5$  (all NCI-H82 groups and Ramos Safe<sup>KO</sup> groups) and 6 (Ramos APMAP<sup>KO</sup> groups)). *P*-values are reported for the interaction between treatment groups. **j**, Single-cell suspensions were prepared from Safe<sup>KO</sup> or APMAP<sup>KO</sup> Ramos tumours treated with PBS or anti-CD20 (from experiment in Fig. [3c](#)) and analysed for the presence of macrophages (CD45<sup>+</sup>/F4-80<sup>+</sup>/Cd11b<sup>+</sup>) as a percentage of all CD45<sup>+</sup> cells. Gating strategy is shown (top/left). Data (bottom right) represent mean  $\pm$  s.e.m. ( $n = 6$  (PBS groups) and 7 (antibody-treated groups)). One-way ANOVA with Tukey correction. **k**, Phagocytosis assay for uptake of pHrodo-labelled B16-F10 cells with indicated genotypes by J774 macrophages in the presence or absence of anti-TRP1 antibodies. Phagocytosis index normalized to control (Safe<sup>KO</sup>) cells without antibody. Data represent mean

$\pm$  s.d. ( $n = 4$  cell culture wells). One-way ANOVA with Bonferroni correction. **I**, In vitro growth of B16-F10 cells of indicated genotypes, measured using time-lapse microscopy as total confluence per well over 6 d. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). **m**, Safe<sup>KO</sup> or APMAP<sup>KO</sup> B16-F10 cells were transplanted into syngeneic C57BL/6 mice and allowed to form tumours. Mice were treated with anti-TRP1 or mouse IgG2a isotype control antibody daily starting 5 d following transplantation, and tumour size was measured every 2 d. Data represent mean  $\pm$  s.e.m. ( $n = 7$  for Safe<sup>KO</sup> groups,  $n = 6$  for both APMAP<sup>KO</sup> groups). Two-way ANOVA with Tukey correction (comparison between Safe<sup>KO</sup>/anti-TRP1 and APMAP<sup>KO</sup>/anti-TRP1 for final timepoint is shown)

[Source data](#).

**Extended Data Fig. 7 Genome-wide magnetic screen in J774 macrophages for phagocytosis of IgG-coated beads.**

**a**, Schematic of genome-wide screen in J774 macrophages for phagocytosis of 2.8 micron IgG-coated magnetic beads. **b**, Volcano plot of screen diagrammed in **a**. Dotted line indicates 5% FDR. **c**, Replicates of screen diagrammed in **a**. **d**, Diagram of hits with negative effect size (i.e. required for phagocytosis) from genome-wide screen for IgG bead phagocytosis in J774 macrophages. **e**, Gene ontology enrichment analysis for macrophage IgG bead screen hits with negative effect size (required for phagocytosis) (5% FDR). Selected terms shown.  $n$  indicates number of genes among hits annotated with indicated term.

**Extended Data Fig. 8 GPR84 is expressed in tumour associated macrophages.**

**a–d**, Single-cell RNA-seq analyses of human tumours from patients with melanoma<sup>54,55</sup> (**a, b**), patients with glioblastoma<sup>56</sup> (**c**) and patients with sarcoma<sup>53</sup> (**d**), showing cell type annotations (left) and detection of GPR84 (right). GPR84+ and GPR84- denote TPM > 0 and = 0, respectively;  $n$  denotes the number of cells shown.

## Extended Data Fig. 9 Macrophage screens for genes required for enhanced uptake of APMAP<sup>KO</sup> cancer cells.

**a**, Phagocytosis assay for uptake of pHrodo-labelled Karpas-299-Cas9 cells expressing indicated sgRNAs, incubated with J774 macrophages expressing indicated sgRNAs, in the presence of anti-CD30 antibodies. Phagocytosis index (arbitrary units) corresponds to the total pHrodo Red signal per well, normalized to Safe<sup>KO</sup> Karpas-299 cells fed to Safe<sup>KO</sup> macrophages. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). Two-way ANOVA with Bonferroni correction. **b**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells expressing indicated sgRNAs, incubated with U937 macrophages expressing indicated sgRNAs (three independent sgRNAs per gene), in the presence of anti-CD20 antibodies. Phagocytosis index (arbitrary units) corresponds to the total pHrodo Red signal per well, normalized to Safe<sup>KO</sup> Ramos cells fed to Safe<sup>KO-1</sup> macrophages. Data represent mean  $\pm$  s.d. ( $n = 4$  cell culture wells). Two-way ANOVA with Bonferroni correction, *P* values for comparisons to Safe<sup>KO</sup> Ramos/Safe<sup>KO-1</sup> U937 macrophages shown. **c**, Gating strategy for collecting single-positive and double-negative macrophage populations, corresponding to macrophages that phagocytosed calcein<sup>+</sup> Safe<sup>KO</sup> or Far-red<sup>+</sup> APMAP<sup>KO</sup> cells. **d**, Volcano plot for macrophage screen for genes required for uptake of Safe<sup>KO</sup> Ramos cells, using 2,208-gene sublibrary (enriched for phagocytosis genes, but lacking GPR84), conducted in J774 macrophages. Dotted line indicates 5% FDR. **e**, Volcano plot for macrophage screen for genes required for uptake of APMAP<sup>KO</sup> Ramos cells, using 2,208-gene sublibrary (enriched for phagocytosis genes, but lacking GPR84) in J774 macrophages. Dotted line indicates 5% FDR. **f**, Volcano plot for macrophage screen for genes required selectively for uptake of APMAP<sup>KO</sup> cells, following screen design in Fig. 4a (comparison 2), but using 2,208-gene sublibrary (enriched for phagocytosis genes, but lacking GPR84) in J774 macrophages, for uptake of calcein<sup>+</sup> Safe<sup>KO</sup> cells and far-red<sup>+</sup> APMAP<sup>KO</sup> Ramos cells. Dotted line indicates 5% FDR. **g**, Phagocytosis assay for uptake of pHrodo-labelled Ramos cells expressing indicated sgRNAs, incubated with J774 macrophages expressing indicated sgRNAs, in the presence of anti-CD20 antibodies. Phagocytosis index (arbitrary units) corresponds to the total pHrodo Red signal per well, normalized to

Safe<sup>KO</sup> Ramos cells fed to Safe<sup>KO</sup> macrophages. Data represent mean +/– s.d. (n = 3 cell culture wells). Two-way ANOVA with Bonferroni correction. P-values correspond to comparisons to Safe<sup>KO-1</sup>. **h**, FACS-based phagocytosis assay for uptake of CellTrace Far-Red-labelled APMAP<sup>KO</sup> cells and calcein-labelled CD47<sup>KO</sup> Ramos cells by J774-Cas9 macrophages expressing indicated sgRNAs. Ratio of macrophages that phagocytosed APMAP<sup>KO</sup> Ramos cells to macrophages that phagocytosed CD47<sup>KO</sup> Ramos cells, normalized to Safe<sup>KO</sup>/Safe<sup>KO</sup> J774 macrophages, following 24 h co-incubation with anti-CD20 antibodies is plotted. Data represent mean +/– s.d. (n = 3 cell culture wells). One-way ANOVA with Bonferroni correction, P-values for comparisons to Safe<sup>KO</sup>/Safe<sup>KO</sup> J774 macrophages shown.

**Extended Data Fig. 10 GPR84 agonists stimulate uptake of antibody-opsonized cancer cells.**

**a**, Phagocytosis assay for uptake of pHrodo-labelled Ramos Cas9 cells expressing Safe-targeting sgRNAs by J774 macrophages in the presence of anti-CD20 antibodies and indicated concentrations of GPR84 agonists. Data represent mean (n = 2 cell culture wells). **b**, Phagocytosis assay for uptake of pHrodo-labelled Ramos Cas9 cells expressing APMAP-targeting sgRNAs by J774 macrophages in the presence of anti-CD20 antibodies and indicated concentrations of GPR84 agonists. Data represent mean (n = 2 cell culture wells). **c**, Phagocytosis assay for uptake of pHrodo-labelled Safe<sup>KO</sup> Ramos Cas9 cells by J774 macrophages in the presence (left) or absence (right) of anti-CD20 antibodies and 100 µM saturated fatty acids of indicated carbon chain length (n = 2, acetic acid; n = 10, capric acid; n = 16, palmitic acid; n = 22, docosanoic acid). Data represent mean +/– s.d. (n = 4 cell culture wells). One-way ANOVA with Bonferroni correction. **d**, Heatmap of normalized phagocytosis index of Ramos cells incubated with U937 macrophages expressing indicated sgRNAs, in the presence of indicated concentrations of 6-OAU and anti-CD20. Data represent mean (n = 4 cell culture wells). **e**, Heatmap of normalized phagocytosis index of Ramos cells incubated with J774 macrophages expressing indicated sgRNAs, in the presence of indicated concentrations of 6-OAU and anti-CD47. Data represent mean (n = 4 cell culture wells). **f**, Phagocytosis assay for uptake of pHrodo-labelled Ramos Cas9 cells expressing Safe-targeting

sgRNAs by J774 macrophages in the presence of anti-CD20 antibodies and GPR84 agonists (100  $\mu$ M capric acid, 100 nM 6-OAU, 10 nM ZQ-16). Data represent mean  $\pm$  s.d. ( $n = 3$  cell culture wells). One-way ANOVA with Bonferroni correction.  $P$ -values are for comparison to untreated condition for each macrophage genotype.

## Supplementary information

### Supplementary Information

This file contains Supplementary Notes 1–5, Supplementary Fig. 1 and full descriptions for Supplementary Tables 1–9.

### Reporting Summary

### Supplementary Table 1

RNA-sequencing data for J774 macrophages with  $P$ -values from two-tailed Wald test, adjusted for multiple comparisons with Benjamini–Hochberg correction.  $n = 3$  biologically independent samples for each condition.

### Supplementary Table 2

Genome-wide ADCP CRISPR knockout screen in Ramos lymphoma cells in presence of anti-CD20.  $P$ -values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5% FDR threshold was used to defining hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicates.

### Supplementary Table 3

Batch re-test ADCP CRISPR knockout screen in Ramos lymphoma cells in presence of anti-CD20. Genes were noted as hits when their combination effect score at 95% credible interval did not include zero. Two biologically independent screen replicates.

## **Supplementary Table 4**

Genome-wide ADCP CRISPR activation screen in Ramos lymphoma cells in presence of anti-CD20 and anti-CD47. *P*-values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5% FDR threshold was used to define hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicates.

## **Supplementary Table 5**

ADCP CRISPR knockout screen in Ramos lymphoma cells in the presence of anti-CD20, +/- anti-CD47, and in sgSafe and sgCD47 genetic backgrounds, using transmembrane protein enriched sublibrary. *P*-values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5% FDR threshold was used to define hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicates in each screen.

## **Supplementary Table 6**

Genome-wide IgG-bead phagocytosis magnetic CRISPR knockout screen in J774 macrophages. *P*-values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5% FDR threshold was used to defining hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicate screens were conducted but one unbound replicate had insufficient coverage so only one unbound replicate was compared to both of the bound replicates.

## **Supplementary Table 7**

Genome-wide ADCP FACS CRISPR knockout screen in J774 macrophages for uptake of Safe<sup>KO</sup> and APMAP<sup>KO</sup> Ramos cells. *P*-values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5%

FDR threshold was used to define hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicates.

## **Supplementary Table 8**

ADCP FACS CRISPR knockout screen in J774 macrophages for uptake of Safe<sup>KO</sup> and APMAP<sup>KO</sup> Ramos cells, using phagocytosis regulator-enriched sublibrary. *P*-values were determined by permuting the gene-targeting guides in the screen and comparing to the distribution of negative controls using casTLE, and a 5% FDR threshold was used to define hits using the Benjamini–Hochberg procedure. Two biologically independent screen replicates.

## **Supplementary Table 9**

sgRNA sequences used in this study.

## **Source data**

### **Source Data Fig. 3**

### **Source Data Extended Data Fig. 6**

## **Rights and permissions**

### **Reprints and Permissions**

## **About this article**

### **Cite this article**

Kamber, R.A., Nishiga, Y., Morton, B. *et al.* Inter-cellular CRISPR screens reveal regulators of cancer cell phagocytosis. *Nature* **597**, 549–554 (2021). <https://doi.org/10.1038/s41586-021-03879-4>

- Received: 14 October 2020
- Accepted: 05 August 2021
- Published: 08 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03879-4>

## 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/s41586-021-03879-4>

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

- Article
- [Published: 08 September 2021](#)

# Using DNA sequencing data to quantify T cell fraction and therapy response

- [Robert Bentham](#)<sup>1,2</sup> na1,
- [Kevin Litchfield](#)<sup>2,3</sup> na1,
- [Thomas B. K. Watkins](#)<sup>4</sup> na1,
- [Emilia L. Lim](#)<sup>2,4</sup>,
- [Rachel Rosenthal](#)<sup>4</sup>,
- [Carlos Martínez-Ruiz](#)<sup>1,2</sup>,
- [Crispin T. Hiley](#)<sup>2,4</sup>,
- [Maise Al Bakir](#)<sup>4</sup>,
- [Roberto Salgado](#) ORCID: orcid.org/0000-0002-1110-3801<sup>5,6</sup>,
- [David A. Moore](#)<sup>2,7,8</sup>,
- [Mariam Jamal-Hanjani](#) ORCID: orcid.org/0000-0003-1212-1259<sup>2,8,9</sup>,
- [TRACERx Consortium](#),
- [Charles Swanton](#) ORCID: orcid.org/0000-0002-4299-3018<sup>2,4,8</sup> &
- [Nicholas McGranahan](#) ORCID: orcid.org/0000-0001-9537-4045<sup>1,2</sup>

[Nature](#) volume 597, pages 555–560 (2021)

- 12k Accesses
- 220 Altmetric
- [Metrics details](#)

## Subjects

- [Biomarkers](#)
- [Immunology](#)
- [Software](#)
- [Tumour immunology](#)

## Abstract

The immune microenvironment influences tumour evolution and can be both prognostic and predict response to immunotherapy<sup>1,2</sup>. However, measurements of tumour infiltrating lymphocytes (TILs) are limited by a shortage of appropriate data. Whole-exome sequencing (WES) of DNA is frequently performed to calculate tumour mutational burden and identify actionable mutations. Here we develop T cell exome TREC tool (T cell ExTRECT), a method for estimation of T cell fraction from WES samples using a signal from T cell receptor excision circle (TREC) loss during V(D)J recombination of the T cell receptor- $\alpha$  gene (*TCRA* (also known as *TRA*)). *TCRA* T cell fraction correlates with orthogonal TIL estimates and is agnostic to sample type. Blood *TCRA* T cell fraction is higher in females than in males and correlates with both tumour immune infiltrate and presence of bacterial sequencing reads. Tumour *TCRA* T cell fraction is prognostic in lung adenocarcinoma. Using a meta-analysis of tumours treated with immunotherapy, we show that tumour *TCRA* T cell fraction predicts immunotherapy response, providing value beyond measuring tumour mutational burden. Applying T cell ExTRECT to a multi-sample pan-cancer cohort reveals a high diversity of the degree of immune infiltration within tumours. Subclonal loss of 12q24.31–32, encompassing *SPPL3*, is associated with reduced *TCRA* T cell fraction. T cell ExTRECT provides a cost-effective technique to characterize immune infiltrate alongside somatic changes.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

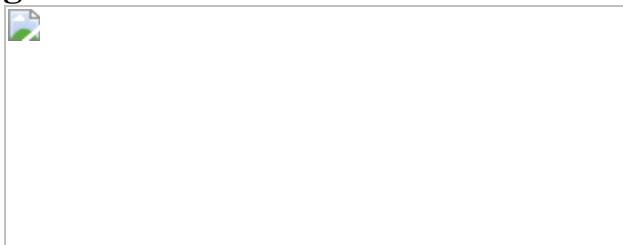
[Rent or Buy](#)

All prices are NET prices.

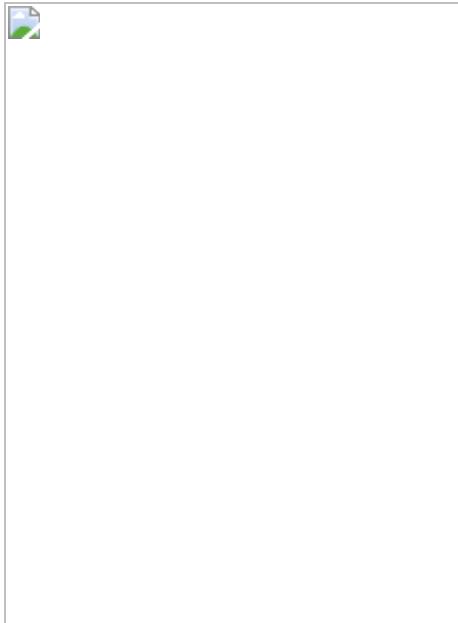
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

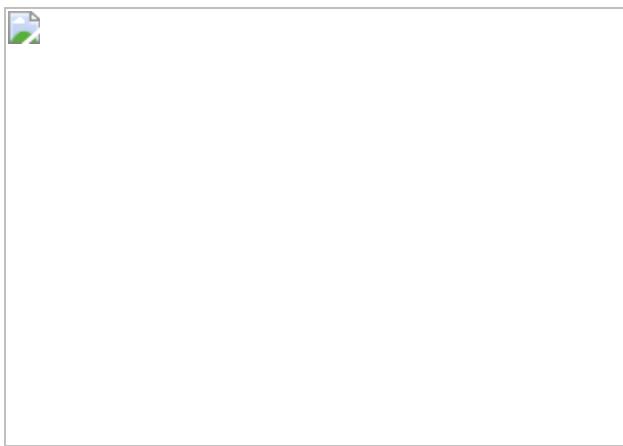
**Fig. 1: Overview and validation of T cell ExTRECT.**



**Fig. 2: Determinants of T cell fraction.**



**Fig. 3: Prognostic value of *TCRA* T cell fraction for LUAD but not for LUSC.**



**Fig. 4: *TCRA* T cell fraction is predictive of survival and response to immunotherapy.**



## Data availability

The RNA-seq data, WES data and histopathology-derived TIL scores (in each case from the TRACERx study) generated, used or analysed during this study are not publicly available and restrictions apply to the availability of these data. Such RNA-seq, WES data and histopathology-derived TIL scores are available through the Cancer Research UK and University College London Cancer Trials Centre ([ctc.tracerx@ucl.ac.uk](mailto:ctc.tracerx@ucl.ac.uk)) for academic non-commercial research purposes upon reasonable request, and subject to review of a project proposal that will be evaluated by a TRACERx data access committee, entering into an appropriate data access agreement and subject to any applicable ethical approvals. Details of all other datasets obtained from third parties used in this study can be found in Extended Data Table 1. Clinical trial information (if applicable) is also available in the associated publications described in Extended Data Table 1.

## Code availability

The code used to produce *TCRA* T cell fraction scores is available for academic non-commercial research purposes at <https://github.com/McGranahanLab/TcellExTRECT>. All other

code used in the analysis and to produce figures is available at  
<https://github.com/McGranahanLab/T-cell-ExTRECT-figure-code-2021>.

## References

1. 1.  
Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
2. 2.  
Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614 (2021).
3. 3.  
Robert, C. et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).
4. 4.  
Schadendorf, D. et al. Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *J. Clin. Oncol.* **33**, 1889–1894 (2015).
5. 5.  
Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **16**, 2598–2608 (2017).
6. 6.  
Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).

7. 7.

Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).

8. 8.

Shen, R. & Seshan, V. *FACETS: Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing, Dept. Epidemiology and Biostatistics Working Paper Series* Vol. 1 No. 50 (Memorial Sloan-Kettering Cancer Center, 2015).

9. 9.

Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

10. 10.

López, S. et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).

11. 11.

Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).

12. 12.

Levy, E. et al. Immune DNA signature of T-cell infiltration in breast tumor exomes. *Sci. Rep.* **6**, 30064 (2016).

13. 13.

Jamal-Hanjani, M. et al. Tracking the evolution of non–small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

14. 14.

Danaher, P. et al. Pan-cancer adaptive immune resistance as defined by the tumor inflammation signature (TIS): results from The Cancer Genome Atlas (TCGA). *J. Immunother. Cancer* **6**, 1–17 (2018).

15. 15.

Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, eaaf8399 (2017).

16. 16.

Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 1–14 (2017).

17. 17.

Li, T. et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* **77**, e108–e110 (2017).

18. 18.

Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

19. 19.

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**, e26476 (2017).

20. 20.

Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).

21. 21.

Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

22. 22.

Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).

23. 23.

Poore, G. D. et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **579**, 567–574 (2020).

24. 24.

Watkins, T. B. K. et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126–132 (2020).

25. 25.

Jongsma, M. L. M. et al. The SPPL3-defined glycosphingolipid repertoire orchestrates HLA class I-mediated immune responses. *Immunity* **54**, 132-150.e9 (2021).

26. 26.

AbdulJabbar, K. et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat. Med.* **26**, 1054–1062 (2020).

27. 27.

Schwartz, L. H. et al. RECIST 1.1—update and clarification: from the RECIST committee. *Eur. J. Cancer* **62**, 132–137 (2016).

28. 28.

Conforti, F. et al. Sex-based dimorphism of anticancer immune response and molecular mechanisms of immune evasion. *Clin. Cancer*

*Res.* **27**, <https://doi.org/10.1158/1078-0432.CCR-21-0136> (2021).

29. 29.

Capone, I., Marchetti, P., Ascierto, P. A., Malorni, W. & Gabriele, L. Sexual dimorphism of immune responses: a new perspective in cancer immunotherapy. *Front. Immunol.* **9**, 552 (2018).

30. 30.

van der Spek, J., Groenwold, R. H. H., van der Burg, M. & van Montfrans, J. M. TREC based newborn screening for severe combined immunodeficiency disease: a systematic review. *J. Clin. Immunol.* **35**, 416–430 (2015).

31. 31.

Kuchenbecker, L. et al. IMSEQ-A fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics* **31**, 2963–2971 (2015).

32. 32.

Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

33. 33.

Middleton, G. et al. The National Lung Matrix Trial of personalized therapy in lung cancer. *Nature* **583**, 807–812 (2020).

34. 34.

Wang, K. et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).

35. 35.

Brastianos, P. K. et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).

36. 36.

Gerlinger, M. et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).

37. 37.

Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).

38. 38.

Harbst, K. et al. Multiregion whole-exome sequencing uncovers the genetic evolution and mutational heterogeneity of early-stage metastatic melanoma. *Cancer Res.* **76**, 4765–4774 (2016).

39. 39.

Lamy, P. et al. Paired exome analysis reveals clonal evolution and potential therapeutic targets in urothelial carcinoma. *Cancer Res.* **76**, 5894–5906 (2016).

40. 40.

Savas, P. et al. The subclonal architecture of metastatic breast cancer: results from a prospective community-based rapid autopsy program “CASCADE”. *PLoS Med.* **13**, e1002204 (2016).

41. 41.

Suzuki, H. et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).

42. 42.

Turajlic, S. et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* **173**, 595-610.e11 (2018).

43. 43.

Messaoudene, M. et al. T-cell bispecific antibodies in node-positive breast cancer: novel therapeutic avenue for MHC class I loss variants. *Ann. Oncol.* **30**, 934–944 (2019).

44. 44.

Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).

45. 45.

Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **352**, 207–212 (2016).

46. 46.

Hugo, W. et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).

47. 47.

Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934-949.e15 (2017).

48. 48.

Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593

(2018).

49. 49.

Snyder, A. et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med.* **14**, 1–24 (2017).

50. 50.

Mariathasan, S. et al. TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).

51. 51.

McDermott, D. F. et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat. Med.* **24**, 749–757 (2018).

52. 52.

Rizvi, N. A. et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).

53. 53.

Le, D. T. et al. PD-1 blockade in tumors with mismatch repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).

54. 54.

Shim, J. H. et al. HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L)1 blockade in advanced non-small-cell lung cancer patients. *Ann. Oncol.* **31**, 902–911 (2020).

55. 55.

Hendry, S. et al. Assessing tumor-infiltrating lymphocytes in solid tumors. *Adv. Anat. Pathol.* **24**, 235–251 (2017).

56. 56.

Denkert, C. et al. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: Results of the ring studies of the international immuno-oncology biomarker working group. *Mod. Pathol.* **29**, 1155–1164 (2016).

## Acknowledgements

R.B. is supported by the NIHR BRC at University College London Hospitals. K.L. is funded by the UK Medical Research Council (MR/P014712/1 and MR/V033077/1), Rosetrees Trust and Cotswold Trust (A2437) and Cancer Research UK (C69256/A30194). T.B.K.W. is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001169), the UK Medical Research Council (FC001169) and the Wellcome Trust (FC001169) as well as the Marie Curie ITN Project PLOIDYNET (FP7-PEOPLE-2013, 607722), Breast Cancer Research Foundation (BCRF), Royal Society Research Professorships Enhancement Award (RP/EA/180007) and the Foulkes Foundation. E.L.L. receives funding from NovoNordisk Foundation (ID 16584). R.R. is supported by Royal Society Research Professorships Enhancement Award (RP/EA/180007). C.M.-R. is supported by Rosetrees. C.T.H. is supported by the NIHR BRC at University College London Hospitals. M.J.-H. has received funding from Cancer Research UK, National Institute for Health Research, Rosetrees Trust, UKI NETs and NIHR University College London Hospitals Biomedical Research Centre. N.M. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant Number 211179/Z/18/Z) and also receives funding from Cancer Research UK, Rosetrees and the NIHR BRC at University College London Hospitals and the CRUK University College London Experimental Cancer Medicine Centre. C.S. is a Royal Society Napier Research Professor. His work was supported by the Francis Crick Institute, which receives its core funding

from Cancer Research UK (FC001169), the UK Medical Research Council (FC001169) and the Wellcome Trust (FC001169). C.S. is funded by Cancer Research UK (TRACERx, PEACE and CRUK Cancer Immunotherapy Catalyst Network), Cancer Research UK Lung Cancer Centre of Excellence, the Rosetrees Trust, Butterfield and Stoneygate Trusts, NovoNordisk Foundation (ID16584), Royal Society Research Professorships Enhancement Award (RP/EA/180007), the NIHR BRC at University College London Hospitals, the CRUK-UCL Centre, Experimental Cancer Medicine Centre and the Breast Cancer Research Foundation (BCRF). This research is supported by a Stand Up To Cancer-LUNGevity-American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (SU2C-AACR-DT23-17). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. C.S. also receives funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) Consolidator Grant (FP7-THESEUS-617844), European Commission ITN (FP7-PloidyNet 607722), an ERC Advanced Grant (PROTEUS) from the European Research Council under the European Union's Horizon 2020 research and innovation programme (835297) and Chromavision from the European Union's Horizon 2020 research and innovation programme (665233). The TRACERx study (Clinicaltrials.gov no: NCT01888601) is sponsored by University College London (UCL/12/0279) and has been approved by an independent Research Ethics Committee (13/LO/1546). TRACERx is funded by Cancer Research UK (C11496/A17786) and coordinated through the Cancer Research UK and UCL Cancer Trials Centre. The results published here are based in part on data generated by The Cancer Genome Atlas pilot project established by the NCI and the National Human Genome Research Institute. The data were retrieved through the database of Genotypes and Phenotypes (dbGaP) authorization (accession number phs000178.v9.p8). Information about TCGA and the constituent investigators and institutions of the TCGA research network can be found at <http://cancergenome.nih.gov/>. This project was enabled through access to the MRC eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council (MR/L016311/1). In particular, we acknowledge the support of the High-Performance Computing at the Francis Crick

Institute as well as the UCL Department of Computer Science Cluster and the support team. In addition, this work was supported by the CRUK City of London Centre Award (C7893/A26233). We thank all investigators, funders and industry partners that supported the generation of the data within this study, as well as patients for their participation. Specifically, we thank Merck & Co, Genentech and Bristol-Myers Squibb for generating the industry datasets used in this study, and E. Van Allen, L. Diaz, T. A. Chan, L. A. Garraway, R. S. Lo, D. F. Bajorin, D. Schadendorf, T. Powles, S.-H. Lee, A. Ribas and S. Ogawa for academic datasets. This work has been funded by Merck and Co. We gratefully acknowledge the patients and their families, the investigators and site personnel who participated in the KEYNOTE-001, -006, -012 and -028 studies.

## Author information

### Author notes

1. These authors contributed equally: Robert Bentham, Kevin Litchfield, Thomas B. K. Watkins

## Affiliations

1. Cancer Genome Evolution Research Group, Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK

Robert Bentham, Carlos Martínez-Ruiz & Nicholas McGranahan

2. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK

Robert Bentham, Kevin Litchfield, Emilia L. Lim, Carlos Martínez-Ruiz, Crispin T. Hiley, David A. Moore, Mariam Jamal-Hanjani, Charles Swanton & Nicholas McGranahan

3. The Tumour Immunogenomics and Immunosurveillance Lab, Cancer Research UK Lung Cancer Centre of Excellence, University College

London Cancer Institute, London, UK

Kevin Litchfield

4. Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK

Thomas B. K. Watkins, Emilia L. Lim, Rachel Rosenthal, Crispin T. Hiley, Maise Al Bakir & Charles Swanton

5. Department of Pathology, GZA-ZNA, Antwerp, Belgium

Roberto Salgado

6. Division of Research, Peter MacCallum Cancer Centre, University of Melbourne, Melbourne, Victoria, Australia

Roberto Salgado

7. Department of Cellular Pathology, University College London Hospitals, London, UK

David A. Moore

8. Department of Medical Oncology, University College London Hospitals, London, UK

David A. Moore, Mariam Jamal-Hanjani & Charles Swanton

9. Cancer Metastasis Lab, University College London Cancer Institute, London, UK

Mariam Jamal-Hanjani

10. The Francis Crick Institute, London, UK

Nicolai J. Birkbak, Mickael Escudero, Aengus Stewart, Andrew Rowan, Jacki Goldman, Peter Van Loo, Richard Kevin Stone, Tamara Denner, Emma Nye, Sophia Ward, Stefan Boeing, Maria

Greco, Jerome Nicod, Clare Puttik, Katey Enfield, Emma Colliver, Brittany Campbell, Alexander M. Frankell, Daniel Cook, Mihaela Angelova, Alastair Magness, Chris Bailey, Antonia Toncheva, Krijn Dijkstra, Judit Kisistok, Mateo Sokac, Oriol Pich, Jonas Demeulemeester, Elizabeth Larose Cadieux, Carla Castignani, Krupa Thakkar, Hongchang Fu, Takahiro Karasaki, Othman Al-Sawaf & Mark S. Hill

11. University College London Cancer Institute, London, UK

Takahiro Karasaki, Othman Al-Sawaf, Christopher Abbosh, Yin Wu, Selvaraju Veeriah, Robert E. Hynds, Andrew Georgiou, Mariana Werner Sunderland, James L. Reading, Sergio A. Quezada, Karl S. Peggs, Teresa Marafioti, John A. Hartley, Helen L. Lowe, Leah Ensell, Victoria Spanswick, Angeliki Karamani, Dhruva Biswas, Stephan Beck, Olga Chervova, Miljana Tanic, Ariana Huebner, Michelle Dietzen, James R. M. Black, Cristina Naceur-Lombardelli, Mita Afroza Akther, Haoran Zhai, Nnennaya Kanu, Simranpreet Summan, Francisco Gimeno-Valiente, Kezhong Chen, Elizabeth Manzano, Supreet Kaur Bola, Ehsan Ghorani, Marc Robert de Massy, Elena Hoxha, Emine Hatipoglu, Benny Chain, David R. Pearce, Javier Herrero & Simone Zaccaria

12. University College London Hospitals, London, UK

Mark S. Hill, David Lawrence, Martin Hayward, Nikolaos Panagiotopoulos, Robert George, Davide Patrini, Mary Falzon, Elaine Borg, Reena Khiroya, Asia Ahmed, Magali Taylor, Junaid Choudhary, Sam M. Janes, Martin Forster, Tanya Ahmad, Siow Ming Lee, Neal Navani, Dionysis Papadatos-Pastos, Marco Scarci, Pat Gorman, Elisa Bertoja, Robert C. M. Stephens, Emilie Martinoni Hoogenboom, James W. Holding, Steve Bandula, Ricky Thakrar, Radhi Anand, Kayalvizhi Selvaraju, James Wilson, Sonya Hessey, Paul Ashford, Mansi Shah & Marcos Vasquez Duran

13. Swansea Bay University Health Board, Swansea, UK

Jason Lester

14. Cardiff and Vale University Health Board, Cardiff, UK

Fiona Morgan, Małgorzata Kornaszewska, Richard Attanoos, Haydn Adams & Helen Davies

15. Cancer Research Centre, University of Leicester, Leicester, UK

Jacqui A. Shaw, Joan Riley, Lindsay Primrose & Dean Fennell

16. Leicester University Hospitals, Leicester, UK

Dean Fennell, Apostolos Nakas, Sridhar Rathinam, Rachel Plummer, Rebecca Boyles, Mohamad Tufail, Amrita Bajaj, Jan Brozik, Keng Ang & Mohammed Fiyaz Chowdhry

17. National Institute for Health Research Leicester Respiratory Biomedical Research Unit, Leicester, UK

William Monteiro & Hilary Marshall

18. University of Leicester, Leicester, UK

Alan Dawson, Sara Busacca, Domenic Marrone & Claire Smith

19. Barnet & Chase Farm Hospitals, Barnet, UK

Girija Anand & Sajid Khan

20. Aberdeen Royal Infirmary, Aberdeen, UK

Gillian Price, Mohammed Khalil, Keith Kerr, Shirley Richardson, Heather Cheyne, Joy Miller, Keith Buchan, Mahendran Chetty & Sylvie Dubois-Marshall

21. The Whittington Hospital NHS Trust, London, UK

Sara Lock & Kayleigh Gilbert

22. University Hospital Birmingham NHS Foundation Trust, Birmingham, UK

Babu Naidu, Gerald Langman, Hollie Bancroft, Salma Kadiri, Gary Middleton, Madava Djearaman, Aya Osman, Helen Shackleford & Akshay Patel

23. Manchester Cancer Research Centre Biobank, Manchester, UK

Angela Leek, Nicola Totten, Jack Davies Hodgkinson, Jane Rogan, Katrina Moore & Rachael Waddington

24. Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK

Raffaele Califano, Rajesh Shah, Piotr Krysiak, Kendadai Rammohan, Eustace Fontaine, Richard Booton, Matthew Evison, Stuart Moss, Juliette Novasio, Leena Joseph, Paul Bishop, Anshuman Chaturvedi, Helen Doran, Felice Granato, Vijay Joshi, Elaine Smith, Angeles Montero & Philip Crosbie

25. Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester, UK

Philip Crosbie

26. Cancer Research UK Lung Cancer Centre of Excellence, University of Manchester, Manchester, UK

Philip Crosbie, Fiona Blackhall, Lynsey Priest, Matthew G. Krebs, Caroline Dive, Dominic G. Rothwell, Alastair Kerr & Elaine Kilgour

27. Christie NHS Foundation Trust, Manchester, United Kingdom

Fiona Blackhall, Lynsey Priest, Matthew G. Krebs, Katie Baker, Mathew Carter, Colin R. Lindsay & Fabio Gomes

28. Cancer Research UK Manchester Institute, University of Manchester, Manchester, UK

Caroline Dive, Dominic G. Rothwell, Alastair Kerr, Elaine Kilgour, Jonathan Tugwood, Jackie Pierce & Alexandra Clipson

29. Berlin Institute for Medical Systems Biology, Max Delbrueck Center for Molecular Medicine, Berlin, Germany

Roland Schwarz & Matthew Huska

30. German Cancer Consortium (DKTK), partner site Berlin, Berlin, Germany

Roland Schwarz

31. Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

Tom L. Kaufmann

32. BIFOLD, Berlin Institute for the Foundations of Learning and Data, Berlin, Germany

Tom L. Kaufmann

33. Danish Cancer Society Research Center, Copenhagen, Denmark

Zoltan Szallasi

34. Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary

Istvan Csabai & Miklos Diossy

35. Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA

Hugo Aerts

36. Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, The Netherlands

Hugo Aerts & Charles Fekete

37. Department of Medical Physics and Bioengineering, University College London Cancer Institute, London, UK

Gary Royle & Catarina Veiga

38. Department of Oncology and Radiotherapy, Medical University of Gdańsk, Gdańsk, Poland

Marcin Skrzypski

39. Independent Cancer Patients Voice, London, UK

Mairead MacKenzie & Maggie Wilcox

40. Cancer Research UK and UCL Cancer Trials Centre, London, UK

Allan Hackshaw, Yenting Ngai, Abigail Sharp, Cristina Rodrigues, Oliver Pressey, Sean Smith, Nicole Gower, Harjot Kaur Dhanda, Kitty Chan & Sonal Chakraborty

41. University Hospital Southampton NHS Foundation Trust, Southampton, UK

Christian Ottensmeier, Serena Chee, Benjamin Johnson, Aiman Alzetani, Judith Cave, Lydia Scarlett & Emily Shaw

42. Royal Brompton and Harefield NHS Foundation Trust, London, UK

Eric Lim, Paulo De Sousa, Simon Jordan, Alexandra Rice, Hilgardt Raubenheimer, Harshil Bhayani, Morag Hamilton, Lyn Ambrose, Anand Devaraj, Hema Chavan, Sofina Begum, Silviu I. Buder, Daniel Kaniu, Mpho Malima, Sarah Booth, Andrew G.

Nicholson, Nadia Fernandes, Christopher Deeley, Pratibha Shah & Chiara Proli

43. Barts Health NHS Trust, London, UK

Kelvin Lau, Michael Sheaff, Peter Schmid, Louise Lim & John Conibear

44. Ashford and St Peter's Hospitals NHS Foundation Trust, Chertsey, UK

Madeleine Hewish

45. Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Sarah Danson, Jonathan Bury, John Edwards, Jennifer Hill, Sue Matthews, Yota Kitsanta, Jagan Rao, Sara Tenconi, Laura Socci, Kim Suvarna, Faith Kibutu, Patricia Fisher, Robin Young, Joann Barker, Fiona Taylor & Kirsty Lloyd

46. Liverpool Heart and Chest Hospital NHS Foundation Trust, Liverpool, UK

Michael Shackcloth & Julius Asante-Siaw

47. Royal Liverpool University Hospital, Liverpool, UK

John Gosney

48. The Princess Alexandra Hospital NHS Trust, Harlow, UK

Teresa Light, Tracey Horey, Peter Russell & Dionysis Papadatos-Pastos

49. NHS Greater Glasgow and Clyde, Glasgow, UK

Kevin G. Blyth, Craig Dick & Andrew Kidd

50. Golden Jubilee National Hospital, Clydebank, UK

Alan Kirk, Mo Asif, John Butler, Rocco Bilancia, Nikos Kostoulas & Mathew Thomas

51. Achilles Therapeutics UK Limited, London, UK

Gareth A. Wilson

## **Consortia**

### **TRACERx Consortium**

- Charles Swanton
- , Mariam Jamal-Hanjani
- , Nicholas McGranahan
- , Carlos Martínez-Ruiz
- , Robert Bentham
- , Kevin Litchfield
- , Emilia L. Lim
- , Crispin T. Hiley
- , David A. Moore
- , Thomas B. K. Watkins
- , Rachel Rosenthal
- , Maise Al Bakir
- , Roberto Salgado
- , Nicolai J. Birkbak
- , Mickael Escudero
- , Aengus Stewart
- , Andrew Rowan
- , Jacki Goldman
- , Peter Van Loo
- , Richard Kevin Stone
- , Tamara Denner
- , Emma Nye
- , Sophia Ward
- , Stefan Boeing
- , Maria Greco
- , Jerome Nicod

- , Clare Puttick
- , Katey Enfield
- , Emma Colliver
- , Brittany Campbell
- , Alexander M. Frankell
- , Daniel Cook
- , Mihaela Angelova
- , Alastair Magness
- , Chris Bailey
- , Antonia Toncheva
- , Krijn Dijkstra
- , Judit Kisistok
- , Mateo Sokac
- , Oriol Pich
- , Jonas Demeulemeester
- , Elizabeth Larose Cadieux
- , Carla Castignani
- , Krupa Thakkar
- , Hongchang Fu
- , Takahiro Karasaki
- , Othman Al-Sawaf
- , Mark S. Hill
- , Christopher Abbosh
- , Yin Wu
- , Selvaraju Veeriah
- , Robert E. Hynds
- , Andrew Georgiou
- , Mariana Werner Sunderland
- , James L. Reading
- , Sergio A. Quezada
- , Karl S. Peggs
- , Teresa Marafioti
- , John A. Hartley
- , Helen L. Lowe
- , Leah Ensell
- , Victoria Spanswick
- , Angeliki Karamani

- , Dhruva Biswas
- , Stephan Beck
- , Olga Chervova
- , Miljana Tanic
- , Ariana Huebner
- , Michelle Dietzen
- , James R. M. Black
- , Cristina Naceur-Lombardelli
- , Mita Afroza Akther
- , Haoran Zhai
- , Nnennaya Kanu
- , Simranpreet Summan
- , Francisco Gimeno-Valiente
- , Kezhong Chen
- , Elizabeth Manzano
- , Supreet Kaur Bola
- , Ehsan Ghorani
- , Marc Robert de Massy
- , Elena Hoxha
- , Emine Hatipoglu
- , Benny Chain
- , David R. Pearce
- , Javier Herrero
- , Simone Zaccaria
- , Jason Lester
- , Fiona Morgan
- , Malgorzata Kornaszewska
- , Richard Attanoos
- , Haydn Adams
- , Helen Davies
- , Jacqui A. Shaw
- , Joan Riley
- , Lindsay Primrose
- , Dean Fennell
- , Apostolos Nakas
- , Sridhar Rathinam
- , Rachel Plummer

- , Rebecca Boyles
- , Mohamad Tufail
- , Amrita Bajaj
- , Jan Brozik
- , Keng Ang
- , Mohammed Fiyaz Chowdhry
- , William Monteiro
- , Hilary Marshall
- , Alan Dawson
- , Sara Busacca
- , Domenic Marrone
- , Claire Smith
- , Girija Anand
- , Sajid Khan
- , Gillian Price
- , Mohammed Khalil
- , Keith Kerr
- , Shirley Richardson
- , Heather Cheyne
- , Joy Miller
- , Keith Buchan
- , Mahendran Chetty
- , Sylvie Dubois-Marshall
- , Sara Lock
- , Kayleigh Gilbert
- , Babu Naidu
- , Gerald Langman
- , Hollie Bancroft
- , Salma Kadiri
- , Gary Middleton
- , Madava Djearaman
- , Aya Osman
- , Helen Shackleford
- , Akshay Patel
- , Angela Leek
- , Nicola Totten
- , Jack Davies Hodgkinson

- , Jane Rogan
- , Katrina Moore
- , Rachael Waddington
- , Jane Rogan
- , Raffaele Califano
- , Rajesh Shah
- , Piotr Krysiak
- , Kendadai Rammohan
- , Eustace Fontaine
- , Richard Booton
- , Matthew Evison
- , Stuart Moss
- , Juliette Novasio
- , Leena Joseph
- , Paul Bishop
- , Anshuman Chaturvedi
- , Helen Doran
- , Felice Granato
- , Vijay Joshi
- , Elaine Smith
- , Angeles Montero
- , Philip Crosbie
- , Fiona Blackhall
- , Lynsey Priest
- , Matthew G. Krebs
- , Caroline Dive
- , Dominic G. Rothwell
- , Alastair Kerr
- , Elaine Kilgour
- , Katie Baker
- , Mathew Carter
- , Colin R. Lindsay
- , Fabio Gomes
- , Jonathan Tugwood
- , Jackie Pierce
- , Alexandra Clipson
- , Roland Schwarz

- , Tom L. Kaufmann
- , Matthew Huska
- , Zoltan Szallasi
- , Istvan Csabai
- , Miklos Diossy
- , Hugo Aerts
- , Charles Fekete
- , Gary Royle
- , Catarina Veiga
- , Marcin Skrzypski
- , David Lawrence
- , Martin Hayward
- , Nikolaos Panagiotopoulos
- , Robert George
- , Davide Patrini
- , Mary Falzon
- , Elaine Borg
- , Reena Khiroya
- , Asia Ahmed
- , Magali Taylor
- , Junaid Choudhary
- , Sam M. Janes
- , Martin Forster
- , Tanya Ahmad
- , Siow Ming Lee
- , Neal Navani
- , Dionysis Papadatos-Pastos
- , Marco Scarci
- , Pat Gorman
- , Elisa Bertoja
- , Robert C. M. Stephens
- , Emilie Martinoni Hoogenboom
- , James W. Holding
- , Steve Bandula
- , Ricky Thakrar
- , Radhi Anand
- , Kayalvizhi Selvaraju

- , James Wilson
- , Sonya Hessey
- , Paul Ashford
- , Mansi Shah
- , Marcos Vasquez Duran
- , Mairead MacKenzie
- , Maggie Wilcox
- , Allan Hackshaw
- , Yenting Ngai
- , Abigail Sharp
- , Cristina Rodrigues
- , Oliver Pressey
- , Sean Smith
- , Nicole Gower
- , Harjot Kaur Dhanda
- , Kitty Chan
- , Sonal Chakraborty
- , Christian Ottensmeier
- , Serena Chee
- , Benjamin Johnson
- , Aiman Alzetani
- , Judith Cave
- , Lydia Scarlett
- , Emily Shaw
- , Eric Lim
- , Paulo De Sousa
- , Simon Jordan
- , Alexandra Rice
- , Hilgardt Raubenheimer
- , Harshil Bhayani
- , Morag Hamilton
- , Lyn Ambrose
- , Anand Devaraj
- , Hema Chavan
- , Sofina Begum
- , Silviu I. Buderu
- , Daniel Kaniu

- , Mpho Malima
- , Sarah Booth
- , Andrew G. Nicholson
- , Nadia Fernandes
- , Christopher Deeley
- , Pratibha Shah
- , Chiara Proli
- , Kelvin Lau
- , Michael Sheaff
- , Peter Schmid
- , Louise Lim
- , John Conibear
- , Madeleine Hewish
- , Sarah Danson
- , Jonathan Bury
- , John Edwards
- , Jennifer Hill
- , Sue Matthews
- , Yota Kitsanta
- , Jagan Rao
- , Sara Tenconi
- , Laura Soccia
- , Kim Suvarna
- , Faith Kibutu
- , Patricia Fisher
- , Robin Young
- , Joann Barker
- , Fiona Taylor
- , Kirsty Lloyd
- , Michael Shackcloth
- , Julius Asante-Siaw
- , John Gosney
- , Teresa Light
- , Tracey Horey
- , Peter Russell
- , Dionysis Papadatos-Pastos
- , Kevin G. Blyth

- , Craig Dick
- , Andrew Kidd
- , Alan Kirk
- , Mo Asif
- , John Butler
- , Rocco Bilancia
- , Nikos Kostoulas
- , Mathew Thomas
- & Gareth A. Wilson

## Contributions

R.B. helped conceive the study, designed and conducted the bioinformatic analysis, and wrote the manuscript. K.L. curated the CPI1000+ cohort used in the study and provided considerable bioinformatic support on its analysis. T.B.K.W. provided considerable bioinformatic support on the analysis of the multi-sample pan-cancer cohort and helped conceive the study and write the manuscript. T.B.K.W. and E.L.L jointly curated the multi-sample pan-cancer cohort used in the study. R.R. and C.M.-R. provided considerable bioinformatic support in the transcriptomic analysis performed in the study, providing RNA-seq immune score metrics and assisting with the RNA-seq gene-expression analysis respectively. R.S., M.A.B., D.A.M. and C.T.H. jointly analysed histopathology-derived TIL estimates. M.J.-H. designed study protocols and helped to analyse patient clinical characteristics. C.S. helped provide study supervision and helped direct the avenues of bioinformatics analysis and also gave feedback on the manuscript. N.M. conceived and supervised the study and helped write the manuscript.

## Corresponding author

Correspondence to Nicholas McGranahan.

## Ethics declarations

## Competing interests

D.A.M. reports speaker fees from AstraZeneca. M.A.B. has consulted for Achilles Therapeutics. R.R. has consulted for and has stock options in Achilles Therapeutics. K.L. has a patent on indel burden and CPI response pending and speaker fees from Roche tissue diagnostics, research funding from CRUK TDL/Ono/LifeArc alliance, and a consulting role with Monopteros Therapeutics. C.T.H. has received speaker fees from AstraZeneca. M.J.-H. is a member of the Scientific Advisory Board and Steering Committee for Achilles Therapeutics. N.M. has stock options in and has consulted for Achilles Therapeutics and holds a European patent in determining HLA LOH (PCT/GB2018/052004). C.S. acknowledges grant support from Pfizer, AstraZeneca, Bristol Myers Squibb, Roche-Ventana, Boehringer-Ingelheim, Archer Dx Inc. (collaboration in minimal residual disease sequencing technologies) and Ono Pharmaceutical; is an AstraZeneca Advisory Board Member and Chief Investigator for the MeRmaiD1 clinical trial; has consulted for Amgen, Pfizer, Novartis, GlaxoSmithKline, MSD, Bristol Myers Squibb, AstraZeneca, Illumina, Genentech, Roche-Ventana, GRAIL, Medicxi, Bicycle Therapeutics, Metabomed and the Sarah Cannon Research Institute; has stock options in Apogen Biotechnologies, Epic Bioscience and GRAIL; and has stock options and is co-founder of Achilles Therapeutics. C.S. holds patents relating to assay technology to detect tumour recurrence (PCT/GB2017/053289); to targeting neoantigens (PCT/EP2016/059401), identifying patent response to immune checkpoint blockade (PCT/EP2016/071471), determining HLA LOH (PCT/GB2018/052004), predicting survival rates of patients with cancer (PCT/GB2020/050221), to treating cancer by targeting Insertion/deletion mutations (PCT/GB2018/051893); identifying insertion/deletion mutation targets (PCT/GB2018/051892); methods for lung cancer detection (PCT/US2017/028013); and identifying responders to cancer treatment (PCT/GB2018/051912).

## Additional information

**Peer review information** *Nature* thanks Florian Markowetz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Overview and validation of T cell ExTRECT.

**a**, Outline of quantification of the *TCRA* T cell fraction utilising V(D)J recombination and TRECs. *top*: Schematic demonstrating how RDR signals are used to detect SCNA gain or loss events in a standard tumour and matched control sample analysis. In this analysis cells consist of three distinct cell types: tumour cells, T cells and all other stromal cells. *bottom*: Schematic of how this same process works when focussing on the *TCRA* gene in relation to V(D)J recombination and TRECs, the lower right panel indicates an increased number of breakpoints detected in the TRACERx100 dataset within the *TCRA* gene relative to surrounding areas of 14q, suggesting that the TREC signal is captured. **b, c**, Plots showing examples of RDR in two TRACERx100 samples demonstrating either increased levels of T cell content in blood compared to matched tumour (**b**) or increased levels of T cell content in tumour compared to matched blood (**c**). VDV segments refer to variable segments in both the TCR $\alpha$  and TCR $\delta$  locus. **d**, *TCRA* T cell fraction (non-GC corrected) value for FFPE and fresh frozen samples for bladder and melanoma tumours within the CPI1000+ cohort (bladder: n = 228, melanoma: n = 297, two sided Wilcoxon rank-sum (Mann-Whitney U) test used, boxplot shows lower quartile, median and upper quartile values). **e**, Summary of linear model for prediction of non-GC corrected *TCRA* T cell fraction from histology and FFPE sample status within the CPI cohort. **f**, Pie charts of calculated *TCRA* T cell fraction from WES of either T cell-derived cell lines or non-T cell derived cell lines, all HCT116 cell lines had calculated fractions < 1 e-15. **g**, Overview of samples in the TRACERx100 cohort. **e**, Association of the CDR3 V(D)J read score based on the iDNA method to *TCRA* T cell fraction in TRACERx100, error bands represent the 95% confidence interval of the fitted linear model.

## Extended Data Fig. 2 Accuracy of *TCRA* T cell fraction by copy number and depth.

**a**, Simulated log RDR from a sample consisting of 24% T cells, 75% tumour, and 1% non-T cell stroma (*TCRA* copy number = 1). **b**, Calculated *TCRA* T cell fraction versus actual T cell fraction value for simulated data **c**, Difference between calculated naive T cell fraction and actual fraction for range of tumour purities and local tumour copy number states at the *TCRA* locus. **d**, Difference between *TCRA* T cell fraction and actual fraction for a range of local tumour copy number to the *TCRA* locus and tumour purities. **e**, Downsampling of 5 TRACERx100 samples to different depths. **f**, Downsampling of simulated data to different depth levels. **g**, Downsampling of the 5 TRACERx100 samples that with the highest CDR3 read counts to different depths and the resulting CDR3 read counts.

## Extended Data Fig. 3 Extended analysis on determinants of *TCRA* T cell fraction.

**a**, Association of blood *TCRA* T cell fraction to histology in TRACERx100 (n = 93 LUAD and LUSC patients). **b**, Predictors of blood *TCRA* T cell fraction in TCGA LUAD and LUSC cohort (left panel: n = 1017, middle panel: n = 976, right panel: n = 714). **c**, Overview of samples in the TCGA LUAD and LUSC cohort. **d**, Summary of mean *TCRA* T cell fraction in PNE cohort. **e**, Overview plot of PNE cohort containing multi-sample microdissected tissue paired with normal blood samples. **f**, Summary of linear model for predicting blood *TCRA* T cell fraction, PNE infiltration defined as *TCRA* T cell fraction > 0.001, ESCC = Oesophageal squamous cell carcinoma, HGD = high grade dysplasia. **g**, Linear model for *TCRA* T cell fraction in PNE samples from genomic factors. **h**, Association of microbial reads from Kraken with *TCRA* T cell fraction in tumour samples (n = 880). **i**, -Log10 p-values for 59 microbial species tested for association with *TCRA* T cell fraction in blood and tumour sample in LUAD and LUSC. Red line represents the significance threshold at P = 0.000423. **j**, The significant hit *Willamsia* in LUAD tumours, red dots represent samples where reads were detected while blue represent samples with no reads detected (n = 501). **k**, The significant hit *Paeniclostridium* in LUSC

tumours (n = 379). All Wilcoxon tests refer to Wilcoxon rank-sum (Mann-Whitney U) tests and are two sided. Boxplots represent lower quartile, median and upper quartile.

### **Extended Data Fig. 4 Subclonal SCNAs and T cell infiltration.**

**a**, Overview of immune heterogeneity across multi-sample pan-cancer cohort with tumour samples ranked by *TCRA* T cell fraction, *upper panel*: histogram of entire cohort, *lower panel*: tumour sample grouped by patients with solid horizontal lines joining regions from the same patient, each line includes 2 or more tumour region and dashed red line is at the mean *TCRA* T cell fraction in the cohort (0.11). **b**, Overview of patients in the multi-sample pan-cancer cohort. **c**, Lower panel: number of tumours in pan-cancer multi-sample cohort with subclonal gains (dark red) or losses (dark blue) across the genome, horizontal lines signify the samples which have more than 30 tumours ([Methods](#)) with subclonal gains or losses. *Upper panel*: - log10(p-value) of the 160 cytoband regions tested for association between *TCRA* T cell fraction and subclonal gains (dark red points) or losses (dark blue points). Red horizontal line marks significance threshold, only one region is significant, a loss event on chromosome 12q24.31-32. **d**, Volcano plot for the RNA-seq analysis in the TRACERx100 cohort between samples with 12q24.31-32 loss and samples without, genes within the locus are labeled, dotted lines at fold change of 0.25 and adjusted P = 0.05.

### **Extended Data Fig 5 Association of *TCRA* T cell fraction with prognosis.**

**a**, Kaplan-Meier curves for the multi-sample TRACERx100 cohort for LUAD (top) and LUSC (bottom) divided by the number of cold samples in the tumour. Immune-hot and immune-cold samples were defined by using the median of all the tumour samples (0.0736) as a threshold. In each Kaplan-Meier curve the included patients were restricted to those with total samples greater than the number of immune-cold samples used in defining the threshold. **b**, Kaplan-Meier curves for overall and progression free survival in the TCGA LUAD cohort, dividing the cohort into immune-hot and immune-cold groups using the mean of the TCGA LUAD cohort (0.109) as a threshold. **c**, Kaplan-Meier curves for the TCGA LUSC, and

TCGA LUAD & LUSC cohorts for overall and progression free survival using the mean of the TCGA LUAD cohort (0.109) as a threshold for distinguishing hot and cold tumours. **d**, Log2(Hazard ratios) from Kaplan-Meier plots for the TCGA separating the tumour samples into immune-hot and immune-cold based on different thresholds from 0 to 0.16 in steps of 0.0025 for overall and progression free survival. **e**, Hazard ratios of separate Cox regression models relating disease free survival to different multi-sample measures related to the *TCRA* T cell fraction in the entire TRACERx100 cohort as well as the LUAD and LUSC patients separately. *TCRA* divergence score is defined as the maximum divided by the upper 95% confidence interval of the minimum. **f**, Hazard ratios of separate Cox regression models for *TCRA* T cell fraction for the TCGA LUAD and LUSC cohort for both overall survival (OS) and progression free survival (PFS).

## Extended Data Fig 6 Overview of CPI1000+ cohort.

**a**, Cohort overview of the CPI1000+ dataset. **b**, Overview of samples in the CPI1000+ cohort excluding Snyder et al.<sup>49</sup> and those with prior CPI treatment. **c**, ROC plot of GLM models for predicting CPI response (blue: clonal TMB, red: clonal TMB + *TCRA* T cell fraction, green: clonal TMB + *CD8A* expression). **d**, Cohort overview of the CPI lung dataset, red lines in upper panel reflect the median *TCRA* T cell fraction in patients with (0.10) or without (0.0070) a response to CPI, note that Tumour *TCRA* T cell fraction particularly in non-responders is often zero. **e**, Overview of patients in the CPI Lung cohort.

## Extended Data Table 1 Original source publications

## Supplementary information

### Supplementary Information

This file contains Supplementary Methods, text, equations, Figs. 1–3 and references

## Reporting Summary

## Rights and permissions

### Reprints and Permissions

## About this article

### Cite this article

Bentham, R., Litchfield, K., Watkins, T.B.K. *et al.* Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* **597**, 555–560 (2021). <https://doi.org/10.1038/s41586-021-03894-5>

- Received: 21 December 2020
- Accepted: 10 August 2021
- Published: 08 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03894-5>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03894-5>

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

- Article
- [Published: 08 September 2021](#)

# Single-cell Ribo-seq reveals cell cycle-dependent translational pausing

- [Michael VanInsberghe](#) ORCID: [orcid.org/0000-0001-8418-4393](https://orcid.org/0000-0001-8418-4393)<sup>1</sup>,
- [Jeroen van den Berg](#)<sup>1</sup>,
- [Amanda Andersson-Rolf](#)<sup>1</sup>,
- [Hans Clevers](#) ORCID: [orcid.org/0000-0002-3077-5582](https://orcid.org/0000-0002-3077-5582)<sup>1</sup> &
- [Alexander van Oudenaarden](#) ORCID: [orcid.org/0000-0002-9442-3551](https://orcid.org/0000-0002-9442-3551)<sup>1</sup>

[Nature](#) volume 597, pages 561–565 (2021)

- 14k Accesses
- 236 Altmetric
- [Metrics details](#)

## Subjects

- [Gene expression profiling](#)
- [RNA sequencing](#)
- [Translation](#)

## Abstract

Single-cell sequencing methods have enabled in-depth analysis of the diversity of cell types and cell states in a wide range of organisms. These tools focus predominantly on sequencing the genomes<sup>1</sup>, epigenomes<sup>2</sup> and transcriptomes<sup>3</sup> of single cells. However, despite recent progress in detecting proteins by mass spectrometry with single-cell resolution<sup>4</sup>, it remains a major challenge to measure translation in individual cells. Here, building on existing protocols<sup>5,6,7</sup>, we have substantially increased the sensitivity of these assays to enable ribosome profiling in single cells. Integrated with a machine learning approach, this technology achieves single-codon resolution. We validate this method by demonstrating that limitation for a particular amino acid causes ribosome pausing at a subset of the codons encoding the amino acid. Of note, this pausing is only observed in a sub-population of cells correlating to its cell cycle state. We further expand on this phenomenon in non-limiting conditions and detect pronounced GAA pausing during mitosis. Finally, we demonstrate the applicability of this technique to rare primary enteroendocrine cells. This technology provides a first step towards determining the contribution of the translational process to the remarkable diversity between seemingly identical cells.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

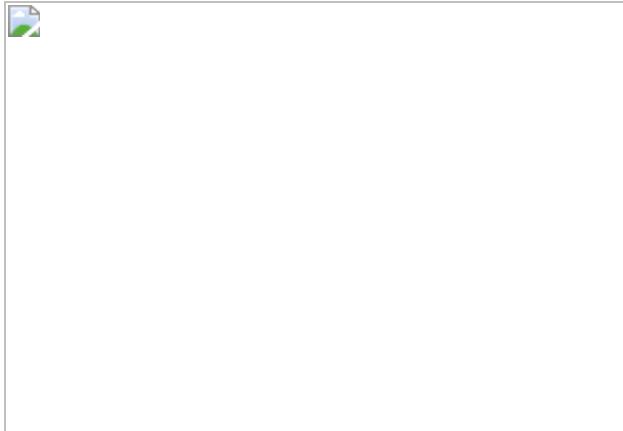
[Rent or Buy](#)

All prices are NET prices.

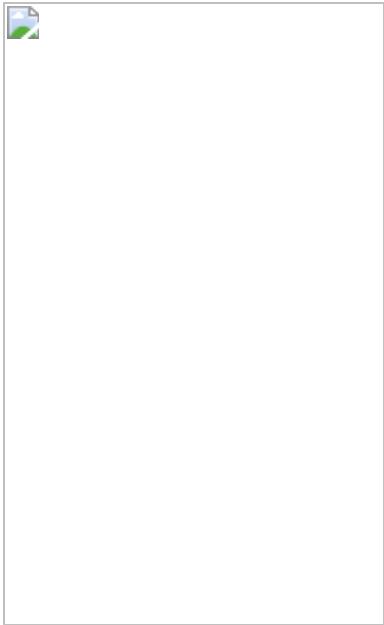
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

**Fig. 1: scRibo-seq measures translation in single cells.**



**Fig. 2: Ribosome pausing under amino acid limitation.**



**Fig. 3: Ribosome pausing during the cell cycle.**



## Data availability

Raw sequencing data, metadata and count tables have been made available in the Gene Expression Omnibus under the accession number [GSE162060](#). Raw sequencing data for comparisons to conventional ribosomal profiling methods were downloaded from Gene Expression Omnibus accessions [GSE37744](#), [GSE125218](#), [GSE113751](#) and [GSE67902](#).

# Code availability

All scripts to process raw data and generate figures are available at [https://github.com/mvanins/scRiboSeq\\_manuscript](https://github.com/mvanins/scRiboSeq_manuscript).

# References

1. 1.

Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).

2. 2.

Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).

3. 3.

Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

4. 4.

Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).

5. 5.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).

6. 6.

Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol. Cell* **71**, 229–243.e11 (2018).

7. 7.

Reid, D. W., Shenolikar, S. & Nicchitta, C. V. Simple and inexpensive ribosome profiling analysis of mRNA translation. *Methods* **91**, 69–74 (2015).

8. 8.

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).

9. 9.

Martinez, T. F. et al. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).

10. 10.

Tanenbaum, M. E., Stern-Ginossar, N., Weissman, J. S. & Vale, R. D. Regulation of mRNA translation during mitosis. *eLife* **4**, e07957 (2015).

11. 11.

Gerashchenko, M. V. & Gladyshev, V. N. Ribonuclease selection for ribosome profiling. *Nucleic Acids Res.* **45**, e6 (2017).

12. 12.

Fang, H. et al. Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Syst.* **6**, 180–191.e4 (2018).

13. 13.

Subramaniam, A. R., Pan, T. & Cluzel, P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl Acad. Sci. USA* **110**, 2419–2424 (2013).

14. 14.

Zinshteyn, B. & Gilbert, W. V. Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* **9**, e1003675 (2013).

15. 15.

Nedialkova, D. D. & Leidel, S. A. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* **161**, 1606–1618 (2015).

16. 16.

Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24**, 2011–2021 (2014).

17. 17.

Stumpf, C. R., Moreno, M. V., Olshen, A. B., Taylor, B. S. & Ruggero, D. The translational landscape of the mammalian cell cycle. *Mol. Cell* **52**, 574–582 (2013).

18. 18.

Coldwell, M. J. et al. Phosphorylation of eIF4GII and 4E-BP1 in response to nocodazole treatment: a reappraisal of translation initiation during mitosis. *Cell Cycle* **12**, 3615–3628 (2013).

19. 19.

Ly, T., Endo, A. & Lamond, A. I. Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *elife* **4**, e04534 (2015).

20. 20.

Miettinen, T. P., Kang, J. H., Yang, L. F. & Manalis, S. R. Mammalian cell growth dynamics in mitosis. *elife* **8**, e44700 (2019).

21. 21.

Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).

22. 22.

Frenkel-Morgenstern, M. et al. Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.* **8**, 572 (2012).

23. 23.

Gribble, F. M. & Reimann, F. Enteroendocrine cells: chemosensors in the intestinal epithelium. *Annu. Rev. Physiol.* **78**, 277–299 (2016).

24. 24.

Gehart, H. et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. *Cell* **176**, 1158–1173.e16 (2019).

25. 25.

Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

26. 26.

Brannan, K. W. et al. Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nat. Methods* **18**, 507–519 (2021).

27. 27.

Shaltiel, I. A. et al. Distinct phosphatases antagonize the p53 response in different phases of the cell cycle. *Proc. Natl Acad. Sci. USA* **111**, 7313–7318 (2014).

28. 28.

Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/060012> (2021).

29. 29.

Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

30. 30.

Ahmed, S., Rattray, M. & Boukouvalas, A. GrandPrix: scaling up the Bayesian GPLVM for single-cell data. *Bioinformatics* **35**, 47–54 (2019).

31. 31.

Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* **19**, 526–541 (2018).

## Acknowledgements

This work was supported by a European Research Council Advanced grant (ERC-AdG 742225-IntScOmics) and Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW 714.016.001). This work is part of the Oncode Institute, which is partly financed by the Dutch Cancer Society. In addition, we thank the Hubrecht

Sorting Facility and the Utrecht Sequencing Facility, subsidized by the University Medical Center Utrecht, the Hubrecht Institute, Utrecht University and The Netherlands X-omics Initiative (NWO project 184.034.019); H. Viñas Gaza for assistance in preparing samples; and V. Bhardwaj for discussion on data analysis.

## Author information

### Affiliations

1. Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Utrecht, The Netherlands

Michael VanInsberghe, Jeroen van den Berg, Amanda Andersson-Rolf, Hans Clevers & Alexander van Oudenaarden

### Contributions

M.V. and A.v.O. conceived and designed the project. M.V. developed the experimental protocol and performed single-cell ribosome profiling experiments with help of J.v.d.B. M.V. and A.v.O analysed the data. M.V., J.v.d.B. and A.v.O. discussed and interpreted results. A.A.-R. and H.C. provided research material. M.V. wrote the manuscript with feedback from A.v.O. and J.v.d.B.

### Corresponding authors

Correspondence to Michael VanInsberghe or Alexander van Oudenaarden.

### Ethics declarations

### Competing interests

The technology described here is the subject of a patent application EP20209743 on which M.V. and A.v.O are inventors.

## Additional information

**Peer review information** *Nature* thanks Arjun Raj, Petra Van Damme and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Library metrics for scRibo-seq libraries.

**a**, Distributions of the number of unique coding-sequence mapped reads per cell. **b**, Distributions of the number of protein-coding genes detected per cell. **c**, Duplicate rate per cell. The mean  $\pm$  standard error of each distribution is indicated.

### Extended Data Fig. 2 Comparison of scRibo-seq to conventional ribosomal profiling.

**a, b**, Heat maps of the percentage of protein-coding reads per library aligning along metagene regions around the start codon (left), in the CDS (middle), and around the stop codon (right). The mapping coordinate of the 5' end (**a**), or the random-forest predicted P-site of each read (**b**) is reported. Libraries are from this work (scRibo-seq), and representative bulk ribosomal profiling methods: Darnell<sup>6</sup>, using MNase on HEK 293T; Ingolia<sup>8</sup>, using RNase I on HEK 293T; Martinez<sup>9</sup>, using RNase I on HEK 293T; and Tanenbaum<sup>10</sup>, using RNase I on RPE-1. **c**, Frame and read-length distributions of the 5' end of RPFs and random-forest predicted P-sites averaged across library sets. **d**, Distributions of the percentage of trimmed reads aligning to rRNA and tRNA. **e**, Region-length normalized distributions of RPF mapping frequencies in the 5' UTR, CDS, and 3' UTR regions of protein-coding transcripts. **f**, Distributions of the percentage of trimmed reads that uniquely align to protein coding, lncRNAs, snoRNAs, or

other biotypes. In the box plots in d-f the middle line indicates the median, the box limits the first and third quartiles, and the whiskers the range. Each point is from a single-cell or bulk library. **g**, Comparisons of the RPF counts per CDS in HEK 293T cells between the different studies. Spearman correlation coefficients for each comparison are indicated.

**Extended Data Fig. 3 A Random Forest model corrects the MNase sequence bias to position ribosome active sites within RPF reads.**

**a**, Logos of the sequence context around the 5' and 3' cut locations. **b**, Schematic illustrating how a nuclease sequence bias can result in a sequence-dependent offset (arrowed lines) between the cut position (triangles) and the ribosome exit, peptidyl, and aminoacyl active sites. Ribosome schematic adapted from ref. <sup>31</sup>. **c**, Schematic describing the parameters used to train the random forest model. Reads spanning a stop codon were used for training. The model predicts the offset between the 5' end of each read and the P-site based on the read length and the sequence context around each end of the read. **d**, Truth table of the model prediction results on validation data. **e**, Permutation importance of the model features. **f**, Frame distributions of the 5' end of RPFs and random-forest predicted P-sites in single cells. Both the 5' and predicted P-sites are uniform between cells and cell types. **g**, Number of footprints per cell along a metagene region within CDS before (top, reads whose 5' ends align at the given region) and after (bottom, number of predicted P-sites at each location) the random forest correction.

**Extended Data Fig. 4 Ribosome pausing in single cells under amino acid limitation.**

**a**, Heat map of the log<sub>2</sub> fold change of amino acid occupancy in the RPF active sites. **b**, Distribution of cells exhibiting ribosome pausing in clusters. The threshold used to distinguish pausing cells was calculated as the mean plus 4 standard deviations of the signal of the cells from the rich condition. **c**, Proportions of treatment type per cluster. **d**, Proportions of treated cells that show a pausing response per cluster. **e**, Gene set enrichment analysis<sup>28</sup>

on the Reactome Pathway database showing the top twenty categories based on marker genes for HEK 293T cell clusters. Categories associated with the cell cycle are highlighted in bold.

**Extended Data Fig. 5 Marker gene expression and site-specific codon abundance over the cell cycle.**

**a**, Heat map of RPF abundance per CDS in hTERT RPE-1 FUCCI cells, showing the translation dynamics of 1,853 significantly differentially translated genes during the cell cycle. Common cell cycle markers are highlighted. **b**, Heat map showing ribosome-site-specific pausing over all codons for hTERT RPE-1 FUCCI cells. Cells are ordered based on cell cycle progression, and codons are clustered based on the average change in the frequency of occurrence across all sites. Codons with significantly different site occupancies between clusters are indicated with an asterisk.

**Extended Data Fig. 6 Ribosome pausing is distinct from changes in codon usage.**

**a**, Frequency of arginine and leucine codons in histone genes compared to all other genes. Histone genes (light grey) are highly enriched in CGC and CGU codons compared to other genes. Histone genes were defined as those in HGNC gene group 864. In the box plots the middle line indicates the median, the box limits the first and third quartiles, and the whiskers the range. Each point represents a gene. **b**, Heat map of the fold change in codon occupancy for CGC and CGU codons in the ribosome active sites (top) and the expression of histone genes (bottom) in RPE-1 cells. The site-agnostic increases in CGC and CGU in RPF active sites are synchronous with the increase in translation of histone genes during late S phase (cluster 5, teal). The increases of CGC and CGU codons in all active sites is distinct from the pattern seen in the GAA site occupancies, where the increase is specific to the A site.

**Extended Data Fig. 7 Scatter plots showing the fold change in gene-wise A-site frequency of occupancy between each cell cluster and the background for the listed codons.**

The increases (GAA, GAG, and AUA) and decreases (CGA) of the A-site abundance affect the majority of the genes detected across clusters.

### Extended Data Fig. 8 Single-cell ribosome profiling in primary mouse intestinal EEC cells.

**a**, UMAP ( $n = 350$  cells) generated using the RPF counts per CDS. Corresponding cell types and associated marker genes for each cluster are indicated. **b**, **c**, UMAPs illustrating the fluorescence of the mNeonGreen (**b**) and dTomato (**c**) markers from the bi-fluorescent Neurog3Chrono reporter<sup>24</sup>. **d**, UMAP depicting the intestinal region origin of each cell. As expected, there is no enrichment of the cell types within each region. **e**, Scatter plots of the Neurog3Chrono fluorescence denoting the position of each cell cluster within the FACS space. As expected, progenitor cells show an increased mNeonGreen fluorescence, that changes through a double-positive population to dTomato-positive as EEC cells develop. **f**, Heat map showing ribosome-site-specific pausing over CAG and GAA codons. To remove any effects of the uneven distribution of RPFs along highly translated hormone genes, any gene that was more than an average of 2.5% of the RPFs per cell was removed from this analysis. **g**, **h**, UMAPs showing the CAG (**g**) and GAA (**h**) pausing. **i**, Heat map showing the distribution of RPF A sites along the *Chgb* CDS. Cells are grouped based on their CAG and GAA pausing status. The position of CAG (orange) and GAA (purple) codons within the CDS are denoted as ticks at the top, with shared prominent pausing sites for each codon indicated with inverted triangles. **j**, **k**, Scatter plots showing the fold change in gene-wise A-site frequency of occurrence between the pausing and non-pausing (normal) cells within each cluster.

### Extended Data Fig. 9 Marker genes and codon pausing for EEC cells.

**a**, Heat map of 1,517 genes significantly differentially expressed between the cell clusters. Common EEC marker genes are indicated. **b**, UMAPs ( $n = 350$  cells) showing the expression of common EEC marker and hormone genes. **c**, Heat map showing ribosome-site-specific pausing for all codons in

the EEC cells. Cells are clustered based on the profiles across the codons. To remove any effects of the uneven distribution of RPFs along highly translated hormone genes, any gene that was more than an average of 2.5% of the RPFs per cell was removed from this analysis (removed genes: *Chga*, *Chgb*, *Clca1*, *Fcgbp*, *Gcg*, *Ghrl*, *Gip*, *Nts*, *Reg4*, *Sst*).

## [Extended Data Fig. 10 Example gating strategies and population frequencies.](#)

**a**, HEK 293T cells. **b–d**, hTERT RPE-1 FUCCI interphase (**b**), contact-inhibition G0 (**c**) and mitotic shake-off fractions (**d**). **e**, Primary mouse EEC cells. Points are pseudocoloured based on density.

## **Supplementary information**

### [Reporting Summary](#)

### [Peer Review File](#)

### [Supplementary Table 1](#)

Sequences of adapters and primers used for library preparation.

## **Rights and permissions**

### [Reprints and Permissions](#)

## **About this article**

### **Cite this article**

VanInsberghe, M., van den Berg, J., Andersson-Rolf, A. *et al.* Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature* **597**, 561–565 (2021). <https://doi.org/10.1038/s41586-021-03887-4>

- Received: 26 November 2020
- Accepted: 06 August 2021
- Published: 08 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03887-4>

## 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/s41586-021-03887-4>

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

- Article
- [Published: 15 September 2021](#)

# Structural basis for tRNA methylthiolation by the radical SAM enzyme MiaB

- [Olga A. Esakova](#) ORCID: [orcid.org/0000-0001-8377-8368<sup>1</sup>](https://orcid.org/0000-0001-8377-8368),
- [Tyler L. Grove](#) ORCID: [orcid.org/0000-0002-4763-0646<sup>2</sup>](https://orcid.org/0000-0002-4763-0646),
- [Neela H. Yennawar](#) ORCID: [orcid.org/0000-0001-7278-659X<sup>3</sup>](https://orcid.org/0000-0001-7278-659X),
- [Arthur J. Arcinas](#) ORCID: [orcid.org/0000-0003-4903-3500<sup>4 nAff6</sup>](https://orcid.org/0000-0003-4903-3500),
- [Bo Wang](#) ORCID: [orcid.org/0000-0002-0381-3686<sup>1</sup>](https://orcid.org/0000-0002-0381-3686),
- [Carsten Krebs](#) ORCID: [orcid.org/0000-0002-3302-7053<sup>1,4</sup>](https://orcid.org/0000-0002-3302-7053),
- [Steven C. Almo](#) ORCID: [orcid.org/0000-0003-2591-5234<sup>2</sup>](https://orcid.org/0000-0003-2591-5234) &
- [Squire J. Booker](#) ORCID: [orcid.org/0000-0002-7211-5937<sup>1,4,5</sup>](https://orcid.org/0000-0002-7211-5937)

*Nature* volume 597, pages 566–570 (2021)

- 2796 Accesses
- 75 Altmetric
- [Metrics details](#)

## Subjects

- [Metalloproteins](#)
- [X-ray crystallography](#)

## Abstract

Numerous post-transcriptional modifications of transfer RNAs have vital roles in translation. The 2-methylthio- $N^6$ -isopentenyladenosine ( $ms^2i^6A$ ) modification occurs at position 37 (A37) in transfer RNAs that contain adenine in position 36 of the anticodon, and serves to promote efficient A:U codon–anticodon base-pairing and to prevent unintended base pairing by near cognates, thus enhancing translational fidelity<sup>1,2,3,4</sup>. The  $ms^2i^6A$  modification is installed onto isopentenyladenosine ( $i^6A$ ) by MiaB, a radical S-adenosylmethionine (SAM) methylthiotransferase. As a radical SAM protein, MiaB contains one  $[Fe_4S_4]_{RS}$  cluster used in the reductive cleavage of SAM to form a 5'-deoxyadenosyl 5'-radical, which is responsible for removing the C<sup>2</sup> hydrogen of the substrate<sup>5</sup>. MiaB also contains an auxiliary  $[Fe_4S_4]_{aux}$  cluster, which has been implicated<sup>6,7,8,9</sup> in sulfur transfer to C<sup>2</sup> of  $i^6A$ 37. How this transfer takes place is largely unknown. Here we present several structures of MiaB from *Bacteroides uniformis*. These structures are consistent with a two-step mechanism, in which one molecule of SAM is first used to methylate a bridging  $\mu$ -sulfido ion of the auxiliary cluster. In the second step, a second SAM molecule is cleaved to a 5'-deoxyadenosyl 5'-radical, which abstracts the C<sup>2</sup> hydrogen of the substrate but only after C<sup>2</sup> has undergone rehybridization from  $sp^2$  to  $sp^3$ . This work advances our understanding of how enzymes functionalize inert C–H bonds with sulfur.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

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

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

[Rent or Buy](#)

All prices are NET prices.

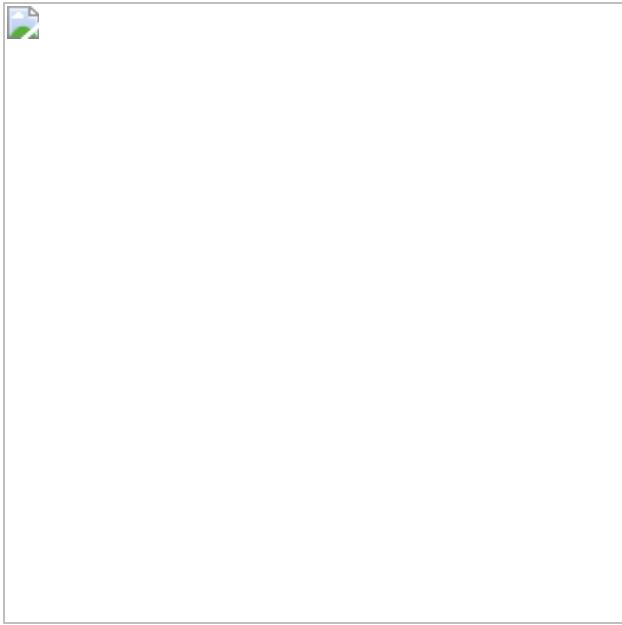
### **Additional access options:**

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

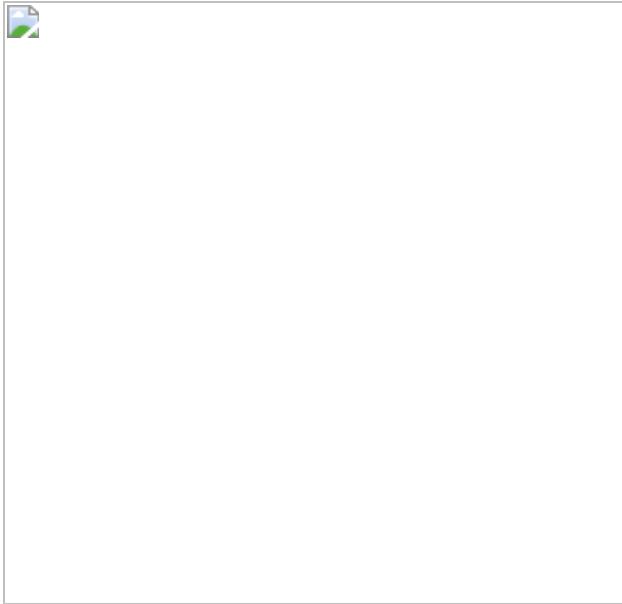
**Fig. 1: Reactions catalysed by MTTases.**



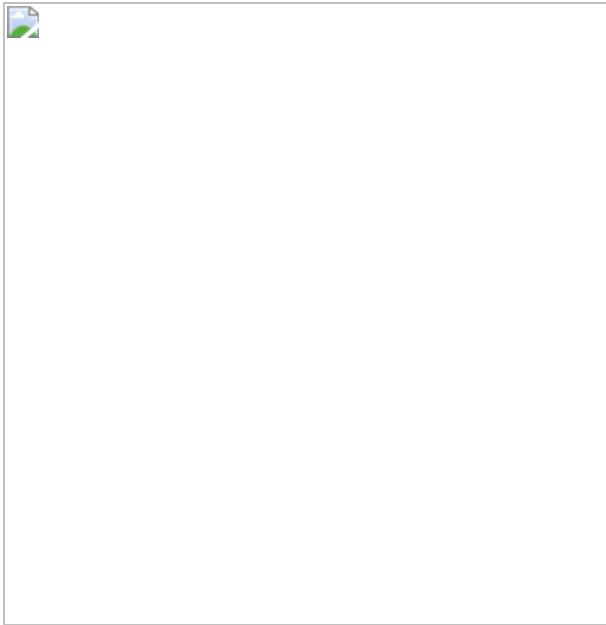
**Fig. 2: RNA binding to *BuMiaB*.**



**Fig. 3: BuMiaB active site in the presence of RNA substrates.**



**Fig. 4: Proposed mechanism for the MiaB reaction.**



## Data availability

Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the PDB under accession numbers [7MJZ](#) (native structure with pentasulfide bridge), [7MJY](#) (structure with SAH and 13-mer RNA), [7MJV](#) (structure with SAM and 17-mer RNA), [7MJX](#) (structure with 5'-dAH+Met and 13-mer RNA) and [7MJW](#) (structure with pre-methylated *BuMiaB* and 5'-dAH+Met and 13-mer RNA).

## References

1. 1.

Connolly, D. M. & Winkler, M. E. Genetic and physiological relationships among the *miaA* gene, 2-methylthio-*N*<sup>6</sup>-(delta 2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in *Escherichia coli* K-12. *J. Bacteriol.* **171**, 3233–3246 (1989).

2. 2.

Connolly, D. M. & Winkler, M. E. Structure of *Escherichia coli* K-12 *miaA* and characterization of the mutator phenotype caused by *miaA* insertion mutations. *J. Bacteriol.* **173**, 1711–1721 (1991).

3. 3.

Esberg, B., Leung, H.-C. E., Tsui, H.-C. T., Björk, G. R. & Winkler, M. E. Identification of the *miaB* gene, involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* **181**, 7256–7265 (1999).

4. 4.

Urbonavicius, J., Qian, Q., Durand, J. M. B., Hagervall, T. G. & Björk, G. R. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **20**, 4863–4873 (2001).

5. 5.

Arcinas, A. J. *Mechanistic Studies of the Radical S-adenosyl-l-methionine (SAM) tRNA Methylthiotransferase MiaB*. Ph.D. thesis, Pennsylvania State Univ., (2016).

6. 6.

Landgraf, B. J., Arcinas, A. J., Lee, K.-H. & Booker, S. J. Identification of an intermediate methyl carrier in the radical S-adenosylmethionine methylthiotransferases RimO and MiaB. *J. Am. Chem. Soc.* **135**, 15404–15416 (2013).

7. 7.

Zhang, B. et al. First step in catalysis of the radical S-adenosylmethionine methylthiotransferase MiaB yields an intermediate with a [3Fe-4S]<sup>0</sup>-like auxiliary cluster. *J. Am. Chem. Soc.* **142**, 1911–1924 (2020).

8. 8.

Forouhar, F. et al. Two Fe-S clusters catalyze sulfur insertion by radical-SAM methylthiotransferases. *Nat. Chem. Biol.* **9**, 333–338 (2013).

9. 9.

Hernández, H. L. et al. MiaB, a bifunctional radical-S-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe–4S] clusters. *Biochemistry* **46**, 5140–5147 (2007).

10. 10.

Reiter, V. et al. The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA. *Nucleic Acids Res.* **40**, 6235–6240 (2012).

11. 11.

Wei, F. Y. et al. Cdk5rap1-mediated 2-methylthio modification of mitochondrial tRNAs governs protein translation and contributes to myopathy in mice and humans. *Cell Metab.* **21**, 428–442 (2015).

12. 12.

Adami, R. & Bottai, D. S-adenosylmethionine tRNA modification: unexpected/unsuspected implications of former/new players. *Int. J. Biol. Sci.* **16**, 3018–3027 (2020).

13. 13.

Dhaven, R. & Tsai, L.-H. A decade of CDK5. *Nat. Rev. Mol. Cell Biol.* **2**, 749–759 (2001).

14. 14.

Arragain, S. et al. Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio- $N^6$ -threonylcarbamoyladenosine in tRNA. *J. Biol. Chem.* **285**, 28425–28433 (2010).

15. 15.

Anton, B. P. et al. Functional characterization of the YmcB and YqeV tRNA methylthiotransferases of *Bacillus subtilis*. *Nucleic Acids Res.* **38**, 6195–6205 (2010).

16. 16.

Landgraf, B. J., McCarthy, E. L. & Booker, S. J. Radical S-adenosylmethionine enzymes in human health and disease. *Annu. Rev. Biochem.* **85**, 485–514 (2016).

17. 17.

Anton, B. P. et al. RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 1826–1831 (2008).

18. 18.

Landgraf, B. J. & Booker, S. J. Stereochemical course of the reaction catalyzed by RimO, a radical SAM methylthiotransferase. *J. Am. Chem. Soc.* **138**, 2889–2892 (2016).

19. 19.

Arragain, S. et al. Post-translational modification of ribosomal proteins: structural and functional characterization of RimO from *Thermotoga maritima*, a radical S-adenosylmethionine methylthiotransferase. *J. Biol. Chem.* **285**, 5792–5801 (2010).

20. 20.

Agris, P. F., Armstrong, D. J., Schäfer, K. P. & Söll, D. Maturation of a hypermodified nucleoside in transfer RNA. *Nucleic Acids Res.* **2**, 691–698 (1975).

21. 21.

Molle, T. et al. Redox behavior of the *S*-adenosylmethionine (SAM)-binding Fe-S cluster in methylthiotransferase RimO, toward understanding dual SAM activity. *Biochemistry* **55**, 5798–5808 (2016).

22. 22.

Lee, T. T., Agarwalla, S. & Stroud, R. M. A unique RNA fold in the RumA–RNA–cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120**, 599–611 (2005).

23. 23.

Anantharaman, V., Koonin, E. V. & Aravind, L. TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol. Lett.* **197**, 215–221 (2001).

24. 24.

Chimnaronk, S. et al. Snapshots of dynamics in synthesizing  $N^6$ -isopentenyladenosine at the tRNA anticodon. *Biochemistry* **48**, 5057–5065 (2009).

25. 25.

Anantharaman, V., Koonin, E. V. & Aravind, L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464 (2002).

26. 26.

Nishimura, S. in *Transfer RNA: Structure, Properties, and Recognition* Vol. 1 (eds Schimmel, P. R., Söll, D. & Abelson, J. R.) (Cold Spring

Harbor Laboratory Press, 1979).

27. 27.

Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).

28. 28.

Ellis, J. J., Broom, M. & Jones, S. Protein–RNA interactions: structural analysis and functional classes. *Proteins* **66**, 903–911 (2007).

29. 29.

Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M. & Thornton, J. M. Protein–RNA interactions: a structural analysis. *Nucleic Acids Res.* **29**, 943–954 (2001).

30. 30.

Pierrel, F., Douki, T., Fontecave, M. & Atta, M. MiaB protein is a bifunctional radical-*S*-adenosylmethionine enzyme involved in thiolation and methylation of tRNA. *J. Biol. Chem.* **279**, 47555–47653 (2004).

31. 31.

Grove, T. L., Radle, M. I., Krebs, C. & Booker, S. J. Cfr and RlmN contain a single [4Fe–4S] cluster, which directs two distinct reactivities for *S*-adenosylmethionine: methyl transfer by  $S_N^2$  displacement and radical generation. *J. Am. Chem. Soc.* **133**, 19586–19589 (2011).

32. 32.

Kim, S., Meehan, T. & Schaefer, H. F., III. Hydrogen-atom abstraction from the adenine–uracil base pair. *J. Phys. Chem. A* **111**, 6806–6812 (2007).

33. 33.

Zierhut, M., Roth, W. & Fischer, I. Dynamics of H-atom loss in adenine. *Phys. Chem. Chem. Phys.* **6**, 5178–5183 (2004).

34. 34.

Lanz, N. D. et al. RlmN and AtsB as models for the overproduction and characterization of radical SAM proteins. *Methods Enzymol.* **516**, 125–152 (2012).

35. 35.

Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn, Vol. 3 (Cold Spring Harbor Laboratory Press, 1989).

36. 36.

McCarthy, E. L. & Booker, S. J. Biochemical approaches for understanding iron–sulfur cluster regeneration in *Escherichia coli* lipoyl synthase during catalysis. *Methods Enzymol.* **606**, 217–239 (2018).

37. 37.

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).

38. 38.

Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).

39. 39.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

40. 40.

Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).

41. 41.

Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

42. 42.

Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

## Acknowledgements

This work was supported by the National Institutes of Health (NIH) (GM-122595 to S.J.B.; AI133329 to S.C.A. and T.L.G.; GM-127079 to C.K.; and GM118393, GM093342 and GM094662 to S.C.A.), the National Science Foundation (MCB-1716686 to S.J.B.), the Eberly Family Distinguished Chair in Science (S.J.B.), The Price Family Foundation (S.C.A.) and the Penn State Huck Institutes of the Life Sciences (N.H.Y.). 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 National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169.

## Author information

### Author notes

1. Arthur J. Arcinas

Present address: AGC Biologics, Seattle, WA, USA

### Affiliations

1. Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

Olga A. Esakova, Bo Wang, Carsten Krebs & Squire J. Booker

2. Department of Biochemistry, Albert Einstein College of Medicine, New York, NY, USA

Tyler L. Grove & Steven C. Almo

3. The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA

Neela H. Yennawar

4. Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

Arthur J. Arcinas, Carsten Krebs & Squire J. Booker

5. The Howard Hughes Medical Institute, The Pennsylvania State University, University Park, PA, USA

Squire J. Booker

## Contributions

O.A.E., T.L.G., N.H.Y. and S.J.B. developed the research plan and experimental strategy. O.A.E. and T.L.G. isolated and crystallized proteins and collected crystallographic data. O.A.E., B.W. and A.J.A. performed biochemical experiments. O.A.E., T.L.G., N.H.Y., S.C.A., C.K. and S.J.B. analysed and interpreted crystallographic data. O.A.E., T.L.G., N.H.Y. and S.J.B. wrote the manuscript, and all other authors reviewed and commented on it.

## Corresponding authors

Correspondence to Olga A. Esakova or Tyler L. Grove or Squire J. Booker.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

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

**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 *BuMiaB* and *TmRimO* structures.

**a**, Cartoon overlay of the structures of *BuMiaB* (blue) and *TmRimO* (PDB:4JC0) (grey). **b**, Electrostatic surface potential of *TmRimO* (blue is positive, red is negative, and grey is neutral). **c**, Amino acid sequence alignment of *BuMiaB* and *TmRimO*. The overall architecture of *BuMiaB* is similar to that of *TmRimO*, with RMSDs of 1.5 and 1.6 Å for the two independent RimO molecules over 324 and 329 C<sub>α</sub>s, respectively (Table S1). In the RimO X-ray crystal structure, the two [Fe<sub>4</sub>S<sub>4</sub>] clusters are 7.3 Å apart (nearest ion in each cluster) and are connected by a pentasulfide bridge spanning the unique (non-cysteinyl-ligated) irons of each cluster. This same pentasulfide bridge is observed in the *BuMiaB* structure, wherein the clusters are 6.8 Å apart (see Extended Data Fig. 2a).

## Extended Data Fig. 2 RNA binds to all three domains in *BuMiaB*.

**a**, Cartoon representation of the active site of *BuMiaB* with a pentasulfide bridge spanning the two [Fe<sub>4</sub>S<sub>4</sub>] clusters. MTTase domain, tan; radical SAM domain, grey; TRAM domain, green. **b**, Cartoon of *BuMiaB* crystallized in the presence of the 13-mer RNA substrate and 5'dAH+Met and showing the binding of the 13-mer (purple) at the interface of the three domains. **c**, Electrostatic surface potential (blue is positive, red is negative, and grey is neutral) indicates a positively charged active-site region promoting binding of the 13-mer. **d**, Conservation of amino acids in the active-site region as deduced from the CONSURF server<sup>42</sup>.

## Extended Data Fig. 3 Comparison of ACSL structure in 13-mer bound to *BuMiaB* with that of full-length tRNA<sup>Phe</sup> bound to *TmMiaA*.

**a**, Cartoon overlay of the 13-mer (nucleotides 29-41) structure in the complex with *BuMiaB* and 5'-dAH+Met (purple colour), with that of *Tm* tRNA<sup>Phe</sup> in complex with MiaA (tan colour) (PDB ID: 2ZM5). **b**,

Schematic diagram of hydrogen bonds formed between the 13-mer and *BuMiaB*.

**Extended Data Fig. 4 Comparison of ACSL structure in 17-mer bound to *BuMiaB* with that of full-length tRNA<sup>Phe</sup> bound to *TmMiaA*.**

**a**, Cartoon overlay of the 13-mer (purple colour) and 17-mer (nucleotides 27-43) (green colour) structures in complex with *BuMiaB* plus 5'-dAH+Met (13-mer) or *BuMiaB* with SAM (17-mer), with that of *Tm* tRNA<sup>Phe</sup> in complex with *MiaA* (tan colour) (**PDB ID: 2ZM5**). **b**, Schematic diagram of the H-bonds formed between the 17-mer and *BuMiaB*.

**Extended Data Fig. 5 Active site interactions of *BuMiaB* with nucleotides 34 and 35 of the anticodon.**

The structure of *BuMiaB* with the 13-mer RNA and 5'dAH+Met is shown in pink, while the structure of *BuMiaB* with the 17-mer RNA and SAM is shown in maroon. **a**, Active site interactions with G34. G34 is often modified, and its base inserts between the TRAM and RS domains in a deep cleft, which provides space for modifications. In the structure of *BuMiaB* in complex with 5'-dAH+Met and the 13-mer, N<sup>10</sup> of G34 is within H-bonding distance to Ser388 of the RS domain. The 2' and 3' OH groups are H-bonded to two nitrogen atoms from the guanidinium group of Arg418 from the TRAM-domain. In the structure of *BuMiaB* with SAM and the 17-mer, the position of G34 is different, and the base no longer interacts with Ser388 and Arg418 (Extended Data Figs. [3b](#), [4b](#)). **b**, Active site interactions with A35. In the structure of *BuMiaB* with 5'-dAH+Met and the 13-mer, the carboxylate oxygens of Asp319 are within H-bonding distance to N<sup>6</sup> of A35, while the side-chain of Gln28 is in H-bonding distance to the 2' OH of A35. The A35 base is π-stacked between Phe348 on one side and the adenine ring of i<sup>6</sup>A37 on the other. The position of A35 is shifted in the SAM-bound structure and is stabilized by π-stacking with G34 on one side and Phe348 on the other. The rotation of Phe348 supports two different orientations of A35 in the active site of the enzyme. **c**, Binding of i<sup>6</sup>A37 in

the active site of *BuMiaB* in the structure with the 13-mer and 5'dAH+Met, showing that the isopentenyl group sits in a hydrophobic patch. All figures have the same colour for the domains and their associated residues: tan for MTTase, grey for radical SAM and green for TRAM.

**Extended Data Fig. 6 A Model for full-length tRNA binding to *BuMiaB*.**

**a**, Conservation of residues in *BuMiaB* as deduced from the CONSURF server. The colour code is described in the panel. **b**, Electrostatic surface potential, indicating positively charged regions that could stabilize tRNA<sup>Phe</sup>. **c**, A predicted model of interactions between the residues from the MTTase domain and the full-length tRNA.

**Extended Data Fig. 7 Binding of SAM, SAH, and 5'-dAH to *BuMiaB*.**

**a**, Overlay of SAM (grey), SAH (aquamarine) and 5'dAH+Met (light violet) in their complexes with *BuMiaB* and RNA substrates (17-mer for SAM, and 13-mer for SAH or 5'-dAH+Met). The adenine ring of SAM, SAH and 5'dAH forms face-to-face  $\pi$ -stacking interactions with Phe321. This stacking is further supported by edge-to-face interactions with two tyrosines (177, 352) and Phe350. N<sup>3</sup> of the adenine ring H-bonds with the conserved Arg66 (shown in Fig. 3a), and N<sup>6</sup> forms three H-bonds with the carbonyl groups of Ile65 (MTTase domain), and Tyr177 and Ser353 (RS domain). The ribose moiety of SAM, SAH and 5'dAH H-bonds with Arg66, Gln281 and Asp319. Methionine in the 5'dAH+Met structure or the methionine moiety of SAM (with the 17-mer RNA) and SAH (with 13-mer RNA) in structures with those molecules bound shows the canonical bidentate binding to the unique iron of the [Fe<sub>4</sub>S<sub>4</sub>]<sub>RS</sub> cluster. **b**, Overlay of SAM (grey) or 5'dAH+Met (light violet) in complex with *BuMiaB* and the 17-mer RNA (SAM) or 13-mer RNA (5'-dAH+Met). The i<sup>6</sup>A37 base is in pink for the structure with 5'dAH+Met and maroon for the structure with SAM. All figures have the same colour for the domains and their associated residues: tan for MTTase, grey for radical SAM and green for TRAM,

except for Gln215 in the structure with 5'dAH+Met (panel a), which rotates.

**Extended Data Fig. 8 Effect of Arg66→Gln substitution on BuMiaB activity.**

**a**, Time course for SAH formation upon incubating 25 μM *BuMiaB* WT (black circles) or *BuMiaB* R66Q (red circles) with 1 mM SAM in the absence of dithionite. **b-e**, Time course for formation of SAH (**b**), 5'dAH (**c**), ms<sup>2</sup>i<sup>6</sup>A (**d**), and decay of i<sup>6</sup>A (**e**), after 30 min of initial incubation of 25 μM *BuMiaB* with 1 mM SAM followed by addition of 100 μM i<sup>6</sup>A ACSL RNA and reaction initiation with 1 mM dithionite. The black colour corresponds to data obtained for *BuMiaB* WT in the presence of the 17-mer RNA substrate; the blue colour corresponds to data obtained in the presence of the 13-mer; and the red colour corresponds to data for *BuMiaB* R66Q in the presence of the 17-mer. Error bars represent one standard deviation for reactions conducted in triplicate, with the centre representing the mean.

**Extended Data Fig. 9 Stereoscopic representation of the active site electron density of pre-methylated *BuMiaB* in the presence of the 13-mer RNA substrate and 5'dAH+Met.**

The structure of *BuMiaB* does not show any significant changes in the protein or RNA components of the complex in the pre-methylated versus non-pre-methylated states [RMSD = 0.231 Å (C<sub>α</sub> = 371 atoms) and 0.089 Å (C<sub>α</sub> = 439 atoms) for pre-methylated subunits A and B, respectively, versus non-pre-methylated subunit A; and 0.092 Å (C<sub>α</sub> = 415 atoms) and 0.248 Å (C<sub>α</sub> = 392 atoms) for pre-methylated subunits A and B, respectively, versus non-methylated subunit B]. **a**, The extended electron density at N<sup>3</sup>. The grey mesh corresponds to an Fo-Fc omit map for i<sup>6</sup>A contoured at 3.5σ, and the green mesh to an Fo-Fc map contoured at 3.0σ after refinement with i<sup>6</sup>A. **b**, The extended electron density at the sulfur atom of the [Fe<sub>3</sub>S<sub>4</sub>] cluster. The mesh corresponds to an Fo-Fc omit map for the methyl group (green colour) attached to the sulfur (grey colour) of the auxiliary cluster contoured at 3.5σ. **c**, In a map generated for the non-pre-methylated

auxiliary cluster, no extended density is observed. The mesh corresponds to an Fo-Fc omit map for the sulfur atom of the auxiliary cluster contoured at  $3.5\sigma$ . All residues have a common colour theme for the domains: tan for MTTase and grey for radical SAM.

## **Extended Data Table 1 X-ray crystallographic data collection and refinement statistics**

# **Supplementary information**

## **Supplementary Information**

This file contains Supplementary Table 1, Supplementary Figs. 1 – 2 and their accompanying legends.

## **Reporting Summary**

# **Rights and permissions**

## **Reprints and Permissions**

# **About this article**

## **Cite this article**

Esakova, O.A., Grove, T.L., Yennawar, N.H. *et al.* Structural basis for tRNA methylthiolation by the radical SAM enzyme MiaB. *Nature* **597**, 566–570 (2021). <https://doi.org/10.1038/s41586-021-03904-6>

- Received: 05 April 2021
- Accepted: 12 August 2021
- Published: 15 September 2021
- Issue Date: 23 September 2021

- DOI: <https://doi.org/10.1038/s41586-021-03904-6>

## 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/s41586-021-03904-6>

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

- Article
- [Published: 08 September 2021](#)

# Positive allosteric mechanisms of adenosine A<sub>1</sub> receptor-mediated analgesia

- [Christopher J. Draper-Joyce](#) ORCID: [orcid.org/0000-0003-2055-7706](https://orcid.org/0000-0003-2055-7706)<sup>1 nAff10</sup>,
- [Rebecca Bhola](#)<sup>2</sup>,
- [Jinan Wang](#) ORCID: [orcid.org/0000-0003-0162-212X](https://orcid.org/0000-0003-0162-212X)<sup>3</sup>,
- [Apurba Bhattacharai](#)<sup>3</sup>,
- [Anh T. N. Nguyen](#)<sup>1</sup>,
- [India Cowie-Kent](#)<sup>2</sup>,
- [Kelly O'Sullivan](#) ORCID: [orcid.org/0000-0003-1715-595X](https://orcid.org/0000-0003-1715-595X)<sup>2</sup>,
- [Ling Yeong Chia](#)<sup>1</sup>,
- [Hariprasad Venugopal](#)<sup>4</sup>,
- [Celine Valant](#)<sup>1</sup>,
- [David M. Thal](#) ORCID: [orcid.org/0000-0002-0325-2524](https://orcid.org/0000-0002-0325-2524)<sup>1</sup>,
- [Denise Wootten](#)<sup>1,5</sup>,
- [Nicolas Panel](#) ORCID: [orcid.org/0000-0001-8782-0586](https://orcid.org/0000-0001-8782-0586)<sup>6</sup>,
- [Jens Carlsson](#)<sup>6</sup>,
- [Macdonald J. Christie](#) ORCID: [orcid.org/0000-0002-0622-609X](https://orcid.org/0000-0002-0622-609X)<sup>7</sup>,
- [Paul J. White](#)<sup>1</sup>,
- [Peter Scammells](#)<sup>8</sup>,
- [Lauren T. May](#)<sup>1</sup>,
- [Patrick M. Sexton](#) ORCID: [orcid.org/0000-0001-8902-2473](https://orcid.org/0000-0001-8902-2473)<sup>1,5</sup>,
- [Radostin Daney](#) ORCID: [orcid.org/0000-0001-6406-8993](https://orcid.org/0000-0001-6406-8993)<sup>9</sup>,
- [Yinglong Miao](#) ORCID: [orcid.org/0000-0003-3714-1395](https://orcid.org/0000-0003-3714-1395)<sup>3</sup>,

- [Alisa Glukhova](#) [ORCID: orcid.org/0000-0003-4146-965X<sup>1</sup>](#) [nAff11](#) [nAff12](#),
- [Wendy L. Imlach](#) [ORCID: orcid.org/0000-0002-7521-9969<sup>2</sup>](#) &
- [Arthur Christopoulos](#) [ORCID: orcid.org/0000-0003-4442-3294<sup>1,5</sup>](#)

[\*Nature\*](#) volume **597**, pages 571–576 (2021)

- 5226 Accesses
- 128 Altmetric
- [Metrics details](#)

## Subjects

- [Cryoelectron microscopy](#)
- [Receptor pharmacology](#)

## Abstract

The adenosine A<sub>1</sub> receptor (A<sub>1</sub>R) is a promising therapeutic target for non-opioid analgesic agents to treat neuropathic pain<sup>1,2</sup>. However, development of analgesic orthosteric A<sub>1</sub>R agonists has failed because of a lack of sufficient on-target selectivity as well as off-tissue adverse effects<sup>3</sup>. Here we show that [2-amino-4-(3,5-bis(trifluoromethyl)phenyl)thiophen-3-yl](4-chlorophenyl)methanone] (MIPS521), a positive allosteric modulator of the A<sub>1</sub>R, exhibits analgesic efficacy in rats *in vivo* through modulation of the increased levels of endogenous adenosine that occur in the spinal cord of rats with neuropathic pain. We also report the structure of the A<sub>1</sub>R co-bound to adenosine, MIPS521 and a G<sub>i2</sub> heterotrimer, revealing an extrahelical lipid–detergent-facing allosteric binding pocket that involves transmembrane helices 1, 6 and 7. Molecular dynamics simulations and ligand kinetic binding experiments support a mechanism whereby MIPS521 stabilizes the adenosine–receptor–G protein complex. This study provides proof of concept for structure-based allosteric drug design of non-opioid analgesic agents that are specific to disease contexts.

## Access options

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Rent or Buy article

Get time limited or full article access on ReadCube.

from \$8.99

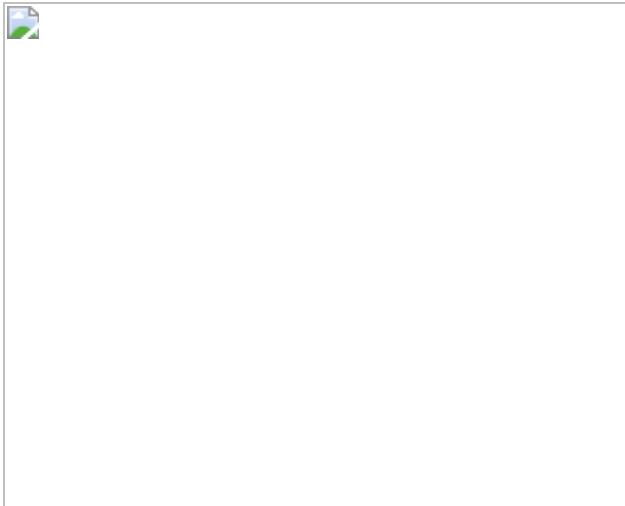
[Rent or Buy](#)

All prices are NET prices.

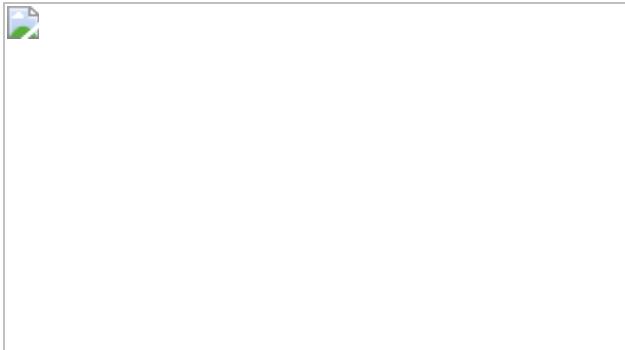
### Additional access options:

- [Log in](#)
- [Access through your institution](#)
- [Learn about institutional subscriptions](#)

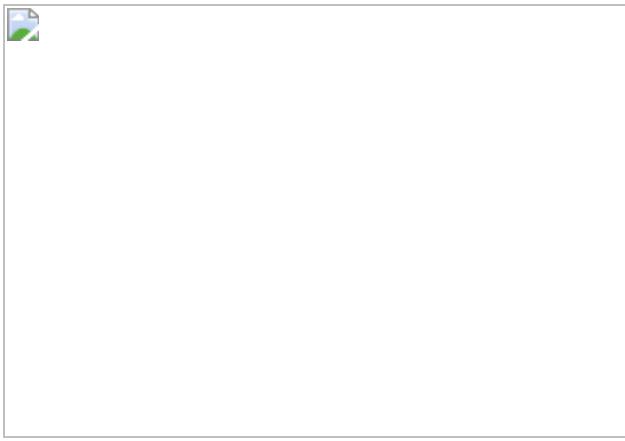
**Fig. 1: MIPS521 reduces spinal nociceptive signalling and mechanical allodynia in an animal model of neuropathic pain.**



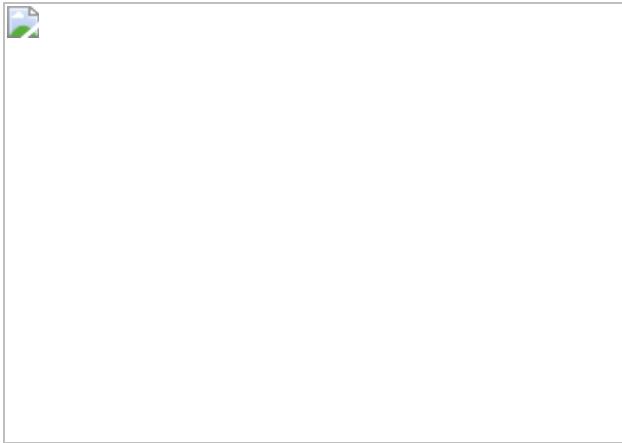
**Fig. 2: Comparison of the structures of the A<sub>1</sub>R–G<sub>i2</sub> complex in the presence and absence of the PAM MIPS521.**



**Fig. 3: Identification of an extrahelical lipid-facing allosteric pocket involving TM1, TM6 and TM7 on the A<sub>1</sub>R.**



**Fig. 4: MIPS521 stabilizes the A<sub>1</sub>R–G<sub>i2</sub> ternary complex.**



## Data availability

Cryo-EM coordinates have been deposited in the PDB under the accession codes [7LD3](#) (MIPS521- and ADO-bound A<sub>1</sub>R–G<sub>i2</sub> complex) and [7LD4](#) (ADO-bound A<sub>1</sub>R–G<sub>i2</sub> complex); the corresponding electron microscopy maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes [EMD-23280](#) and [EMD-23281](#).

## References

1. 1.

Nakamura, I., Ohta, Y. & Kemmotsu, O. M. Characterization of adenosine receptors mediating spinal sensory transmission related to nociceptive information in the rat. *Anesthesiology* **87**, 577–584 (1997).

2. 2.

Poon, A. & Sawynok, J. Antinociception by adenosine analogs and inhibitors of adenosine metabolism in an inflammatory thermal hyperalgesia model in the rat. *Pain* **74**, 235–245 (1998).

3. 3.

Zylka, M. J. Pain-relieving prospects for adenosine receptors and ectonucleotidases. *Trends Mol. Med.* **17**, 188–196 (2011).

4. 4.

King, A. Analgesia without opioids. *Nature* **573**, S4 (2019).

5. 5.

Busse, J. W. et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. *JAMA* **320**, 2448–2460 (2018).

6. 6.

Ribeiro, J. A., Sebastião, A. M. & de Mendonça, A. Adenosine receptors in the nervous system: pathophysiological implications. *Prog. Neurobiol.* **68**, 377–392 (2002).

7. 7.

Choca, J. I., Proudfoot, H. K. & Green, R. D. Identification of A<sub>1</sub> and A<sub>2</sub> adenosine receptors in the rat spinal cord. *J. Pharmacol. Exp. Ther.* **242**, 905–910 (1987).

8. 8.

Choca, J. I., Green, R. D. & Proudfoot, H. K. Adenosine A<sub>1</sub> and A<sub>2</sub> receptors of the substantia gelatinosa are located predominantly on intrinsic neurons: an autoradiography study. *J. Pharmacol. Exp. Ther.* **247**, 757–764 (1988).

9. 9.

Yang, Z. et al. Cardiac overexpression of A<sub>1</sub>-adenosine receptor protects intact mice against myocardial infarction. *Am. J. Physiol. Heart. Circ. Physiol.* **282**, H949–H955 (2002).

10. 10.

Christopoulos, A. & Kenakin, T. G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev.* **54**, 323–374 (2002).

11. 11.

May, L. T. et al. Allosteric modulation of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **47**, 1–51 (2007).

12. 12.

Bruns, R. F. & Fergus, J. H. Allosteric enhancement of adenosine A<sub>1</sub> receptor binding and function by 2-amino-3-benzoylthiophenes. *Mol. Pharmacol.* **38**, 939–949 (1990).

13. 13.

Li, X. et al. Spinal noradrenergic activation mediates allodynia reduction from an allosteric adenosine modulator in a rat model of neuropathic pain. *Pain* **97**, 117–125 (2002).

14. 14.

Childers, S. R. et al. Allosteric modulation of adenosine A<sub>1</sub> receptor coupling to G-proteins in brain. *J. Neurochem.* **93**, 715–723 (2005).

15. 15.

Vincenzi, F. et al. TRR469, a potent A<sub>1</sub> adenosine receptor allosteric modulator, exhibits anti-nociceptive properties in acute and neuropathic pain models in mice. *Neuropharmacology* **81**, 6–14 (2014).

16. 16.

Gramec, D., Mašič, L. P. & Dolenc, M. S. Bioactivation potential of thiophene-containing drugs. *Chem. Res. in Toxicol.* **27**, 1344–1358 (2014).

17. 17.

Nguyen, A. T. et al. Role of the second extracellular loop of the adenosine A<sub>1</sub> receptor on allosteric modulator binding, signaling, and cooperativity. *Mol. Pharmacol.* **90**, 715–725 (2016).

18. 18.

Miao, Y. et al. Structural basis for binding of allosteric drug leads in the adenosine A<sub>1</sub> receptor. *Sci. Rep.* **8**, 16836 (2018).

19. 19.

Glukhova, A. et al. Structure of the adenosine A<sub>1</sub> receptor reveals the basis for subtype selectivity. *Cell* **168**, 867–877 (2017).

20. 20.

Thal, D. M. et al. Structural insights into G-protein-coupled receptor allostery. *Nature* **559**, 45–53 (2018).

21. 21.

Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* **11**, 3584–3595 (2015).

22. 22.

Imlach, W. L. et al. A positive allosteric modulator of the adenosine A<sub>1</sub> receptor selectively inhibits primary afferent synaptic transmission in a neuropathic pain model. *Mol. Pharmacol.* **88**, 460–468 (2015).

23. 23.

Aurelio, L. et al. Allosteric modulators of the adenosine A<sub>1</sub> receptor: synthesis and pharmacological evaluation of 4-substituted 2-amino-3-benzoylthiophenes. *J. Med. Chem.* **52**, 4543–4547 (2009).

24. 24.

Valant, C. et al. Separation of on-target efficacy from adverse effects through rational design of a bitopic adenosine receptor agonist. *Proc. Natl Acad. Sci. USA* **111**, 4614–4619 (2014).

25. 25.

Schulte, G. et al. Distribution of antinociceptive adenosine A<sub>1</sub> receptors in the spinal cord dorsal horn, and relationship to primary afferents and neuronal subpopulations. *Neuroscience* **121**, 907–916 (2003).

26. 26.

Wu, Z.-Y. et al. Endomorphin-2 decreases excitatory synaptic transmission in the spinal ventral horn of the rat. *Front. Neural Circuits* **11**, 55–55 (2017).

27. 27.

Geiger, J. G., LaBella, F. S. & Nagy, J. I. Characterization and localization of adenosine receptors in rat spinal cord. *J. Neurosci.* **4**, 2303–2310 (1984).

28. 28.

Johansson, B. et al. Hyperalgesia, anxiety, and decreased hypoxic neuroprotection in mice lacking the adenosine A<sub>1</sub> receptor. *Proc. Natl Acad. Sci. USA* **98**, 9407–9412 (2001).

29. 29.

Liang, Y. L. et al. Dominant negative G proteins enhance formation and purification of agonist–GPCR–G protein complexes for structure determination. *ACS Pharmacol. Transl. Sci.* **1**, 12–20 (2018).

30. 30.

Draper-Joyce, C. J. et al. Structure of the adenosine-bound human adenosine A<sub>1</sub> receptor–G<sub>i</sub> complex. *Nature* **558**, 559–563 (2018).

31. 31.

Leach, K. et al. Molecular mechanisms of action and in vivo validation of an M4 muscarinic acetylcholine receptor allosteric modulator with potential antipsychotic properties. *Neuropharmacology* **35**, 855–869 (2010).

32. 32.

Zhang, D. et al. Two disparate ligand-binding sites in the human P2Y<sub>1</sub> receptor. *Nature* **520**, 317–321 (2015).

33. 33.

Cheng, R. K. et al. Structural insight into allosteric modulation of protease-activated receptor 2. *Nature* **545**, 112–115 (2017).

34. 34.

Robertson, N. et al. Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature* **553**, 111–114 (2018).

35. 35.

Shao, Z. et al. Structure of an allosteric modulator bound to the CB1 cannabinoid receptor. *Nat. Chem. Biol.* **15**, 1199–1205 (2019).

36. 36.

Lu, J. et al. Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat. Struct. Mol. Biol.* **24**, 570–577 (2017).

37. 37.

Liu, X. et al. Mechanism of  $\beta_2$ AR regulation by an intracellular positive allosteric modulator. *Science* **364**, 1283–1287 (2019).

38. 38.

Zhuang, Y. et al. Mechanism of dopamine binding and allosteric modulation of the human D1 dopamine receptor. *Cell Res.* **31**, 593–596 (2021).

39. 39.

DeVree, B. T. et al. Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature* **535**, 182–186 (2016).

40. 40.

Seltzer, Z., Dubner, R. & Shir, Y. A novel behavioral model of neuropathic pain disorders produced in rats by partial sciatic nerve injury. *Pain* **43**, 205–218 (1990).

41. 41.

Imlach, W. L. et al. Glycinergic dysfunction in a subpopulation of dorsal horn interneurons in a rat model of neuropathic pain. *Sci. Rep.* **6**, 37104 (2016).

42. 42.

Bonin, R. P., Bories, C. & De Koninck, Y. A simplified up-down method (SUDO) for measuring mechanical nociception in rodents using von Frey filaments. *Mol. Pain* **10**, 10–26 (2014).

43. 43.

Størkson, R. V. et al. Lumbar catheterization of the spinal subarachnoid space in the rat. *J. Neurosci. Methods* **65**, 167–172 (1996).

44. 44.

King, T. et al. Unmasking the tonic-aversive state in neuropathic pain. *Nat. Neurosci.* **12**, 1364–1366 (2009).

45. 45.

Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).

46. 46.

Schorb, M. et al. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).

47. 47.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

48. 48.

Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

49. 49.

Maeda, S. et al. Development of an antibody fragment that stabilizes GPCR/G-protein complexes. *Nat. Commun.* **9**, 3712 (2018).

50. 50.

Emsley, P. et al. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

51. 51.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–

221 (2010).

52. 52.

Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

53. 53.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

54. 54.

Baltos, J. A. et al. Quantification of adenosine A<sub>1</sub> receptor biased agonism: implications for drug discovery. *Biochem. Pharmacol.* **99**, 101–112 (2016).

55. 55.

Nguyen, A. T. et al. Extracellular loop 2 of the adenosine A<sub>1</sub> receptor has a key role in orthosteric ligand affinity and agonist efficacy. *Mol. Pharmacol.* **90**, 703–714 (2016).

56. 56.

Morris, G. M. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

57. 57.

Dror, R. O. et al. Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* **348**, 1361–1365 (2015).

58. 58.

Dror, R. O. et al. Activation mechanism of the  $\beta_2$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 18684–18689 (2011).

59. 59.

Wang, J. & Miao, Y. Mechanistic insights into specific G protein interactions with adenosine receptors. *J. Phys. Chem. B* **123**, 6462–6473 (2019).

60. 60.

Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

61. 61.

Vanommeslaeghe, K. & MacKerell, A. D. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* **1850**, 861–871 (2015).

62. 62.

Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).

63. 63.

Klauda, J. B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).

64. 64.

Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).

65. 65.

Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).

66. 66.

Miao, Y. & McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. *Proc. Natl Acad. Sci. USA* **113**, 12162–12167 (2016).

67. 67.

Miao, Y. & McCammon, J. A. Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **115**, 3036–3041 (2018).

68. 68.

Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

69. 69.

Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089 (1993).

70. 70.

Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

71. 71.

Bernstein, N. et al. QM/MM simulation of liquid water with an adaptive quantum region. *Phys. Chem. Chem. Phys.* **14**, 646–656

(2012).

72. 72.

Roe, D. R. & Cheatham, T. E. PTraj and CPPtraj: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

73. 73.

Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).

## Acknowledgements

This work was supported by the National Health and Medical Research Council of Australia (NHMRC) project grants 1145420 and 1147291, NHMRC program grant 1050083, American Heart Association grant 17SDG33370094 and the National Institutes of Health grant R01GM132572. P.M.S., W.L.I. and D.W. are NHMRC Senior Principal Research, Career Development and Senior Research Fellows, respectively. C.J.D.-J., A.G. and D.M.T. are Australian Research Council Discovery Early Career Research Fellows. L.T.M. is an Australian Heart Foundation Future Leaders Fellow. J.C. acknowledges support from the Swedish Research Council (2017-04676). We are grateful to S. Charman and K. White for performing the VCP171 and MIPS521 plasma and liver microsome stability studies. We acknowledge use of facilities within the Monash Ramaciotti Cryo-EM platform. This work was supported by the MASSIVE HPC facility (<https://www.massive.org.au>) and the Extreme Science and Engineering Discovery Environment supercomputing award TG-MCB180049.

## Author information

Author notes

**1. Christopher J. Draper-Joyce**

Present address: The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia

**2. Alisa Glukhova**

Present address: Structural Biology Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

**3. Alisa Glukhova**

Present address: Department of Biochemistry and Pharmacology, University of Melbourne, Parkville, Victoria, Australia

## **Affiliations**

**1. Drug Discovery Biology and Department of Pharmacology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia**

Christopher J. Draper-Joyce, Anh T. N. Nguyen, Ling Yeong Chia, Celine Valant, David M. Thal, Denise Wootten, Paul J. White, Lauren T. May, Patrick M. Sexton, Alisa Glukhova & Arthur Christopoulos

**2. Department of Physiology, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia**

Rebecca Bhola, India Cowie-Kent, Kelly O'Sullivan & Wendy L. Imlach

**3. Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, KS, USA**

Jinan Wang, Apurba Bhattacharai & Yinglong Miao

4. Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia

Hariprasad Venugopal

5. ARC Centre for Cryo-electron Microscopy of Membrane Proteins, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia

Denise Wootten, Patrick M. Sexton & Arthur Christopoulos

6. Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

Nicolas Panel & Jens Carlsson

7. Discipline of Pharmacology, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, Australia

Macdonald J. Christie

8. Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia

Peter Scammells

9. Graduate School of Medicine, University of Tokyo, Tokyo, Japan

Radostin Danev

## Contributions

C.J.D.-J. developed the expression and purification strategy, performed virus production, insect cell expression and purification, generated nanodisc and nanodisc-based pharmacological assays, performed negative-stain EM data acquisition and analysis, and prepared samples for cryo-EM. R.B. prepared pain models for *in vivo* studies, surgical placement of intrathecal catheters, drug administration and behavioural testing (von Frey and

rotarod), and analysed in vivo data. K.O. prepared pain models for electrophysiology studies and pain behavioural testing (von Frey) on rats used for electrophysiology. I.C.-K. assisted R.B. with behavioural assays. W.L.I. performed spinal cord electrophysiology, surgeries for pain models and intrathecal catheter placement, evoked pain behaviour (von Frey) and spontaneous pain behaviour (conditioned place preference) studies, supervised experiments and oversaw experimental design of ex vivo and in vivo experiments. L.Y.C conducted atrial contraction organ bath experiments. P.J.W. oversaw atrial contraction design, experiments and analysis. J.W. and A.B. designed, performed and analysed molecular dynamics simulations. Y.M. oversaw molecular dynamics simulations and analysis. N.P. and J.C. designed, performed and analysed molecular docking studies. D.M.T. developed the expression and purification strategy and assisted with biochemistry and reconstitution of nanodiscs. H.V. organized microscopy time and provided oversight of image acquisition within the Monash EM facility. A.T.N.N. performed whole cell radioligand binding pharmacological assays. A.T.N.N. performed cAMP pharmacological assays, designed the A<sub>1</sub>R mutation strategy, generated mutant A<sub>1</sub>Rs and associated stable cell lines, and performed whole-cell radioligand binding pharmacological assays. L.T.M. supervised A<sub>1</sub>R mutagenesis, whole-cell pharmacological assays and atrial contraction assays. C.J.D.-J., A.T.T.N., C.V. and L.T.M. performed data analysis. P.S. supervised medicinal chemistry design and synthesis. R.D. performed sample plunging for cryo-EM, imaging and data collection. R.D., M.J.C., L.T.M., D.W. and P.M.S. assisted with data interpretation and preparation of the manuscript. A.G. developed the expression and purification strategy, performed negative stain transmission EM, cryo-EM data processing, model building, refinement and validation. C.J.D.-J., Y.M., W.L.I., A.G. and A.C. wrote the manuscript. P.M.S., Y.M., A.G., W.L.I. and A.C. supervised the project.

## Corresponding authors

Correspondence to Alisa Glukhova or Wendy L. Imlach or Arthur Christopoulos.

## Ethics declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature* thanks Grégory Scherrer, Irina Tikhonova, Daniel Wacker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### [Extended Data Fig. 1 Physiological effects of VCP171 and MIPS521.](#)

**a**, Chemical structure of VCP171. **b**, Time courses of paw withdrawal threshold (PWT) to mechanical stimulus by von Frey filaments in nerve-injured rats post-intrathecal administration of VCP171 (blue) or MIPS521 (red). Significance to vehicle control was determined using Greenhouse-Geisser correction for multiple comparisons, corrected with Dunnett's post-hoc test, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ . Data are shown as mean +/- SEM (n=8-10 rats per data group) **c**, Single trial place preference conditioning with intrathecal VCP171 (30 µg, blue), MIPS521 (10 µg, red) and morphine (10 µg, black) increased the time nerve-injured rats spent in the drug paired chamber, with a corresponding decrease in the vehicle paired chamber. Sham surgery rats showed no chamber preference. Empty circles show individual data points, and bars show mean +SEM (n = 8 per group). Significance was determined using a two-tailed unpaired *t* test assuming unequal variance, \*  $P < 0.05$ , \*\*  $P < 0.01$ , compared to vehicle control. **d**, Rotarod latency in rats following intrathecal administration of VCP171 (blue) or MIPS521 (red) is not significantly different to vehicle controls, whereas intrathecal administration of morphine reduces rotarod latency to fall. Data are shown as mean +/- SEM ( $n = 3-4$  per group).

Significance was determined using a two-tailed unpaired *t* test assuming unequal variance, \*  $P < 0.05$ , \*\*  $P < 0.01$ , compared to vehicle control. **e**, Effect of CPA (black;  $n = 4$ ) or MIPS521 (solid red;  $n = 6$ ) on rate of atrial contraction. Data represent mean  $\pm$  SD.

## Extended Data Fig. 2

Effects of VCP171 and MIPS521 on spontaneous excitatory synaptic activity. **a**, Examples of spontaneous excitatory postsynaptic potentials (sEPSCs) recorded from neurons of the superficial laminae of the spinal dorsal horn of nerve-injured rats. **b**, sEPSC frequency and amplitude were reduced following superfusion of VCP171 or MIPS521, which is reversed by the antagonist, DPCPX ( $n = 8$  per group). Significance compared to baseline was determined using a two-tailed paired *t* test, \* $P < 0.05$ , \*\* $P < 0.01$ .

## Extended Data Fig. 3 Expression and purification of the MIPS521–ADO–A<sub>1</sub>R–G<sub>i2</sub> complex.

**a**, Expression and purification flowchart for the A<sub>1</sub>R–G<sub>i2</sub> complex. A<sub>1</sub>R and the G<sub>i2</sub> heterotrimer with G $\beta_1\gamma_2$  were expressed separately in insect cell membranes. Addition of ADO (1 mM) and MIP521 (100 nM) initiated complex formation, which was solubilized with 0.5% (w/v) lauryl maltose neopentyl glycol and 0.05% (w/v) cholesteryl hemisuccinate. Solubilized A<sub>1</sub>R and A<sub>1</sub>R–G<sub>i2</sub> complex was immobilized on Flag antibody resin. Flag-eluted fractions were purified by size-exclusion chromatography (SEC). Illustrations taken from ChemDraw. **b**, SDS–PAGE/western blot of the purified A<sub>1</sub>R–G<sub>i2</sub> complex. An anti-His antibody was used to detect Flag–A<sub>1</sub>R-His and G $\beta_1$ -His (red) and an anti-G<sub>i2</sub> antibody was used to detect G $\alpha_{i2}$  (green). For gel source data, see Supplementary Fig. 1. **c**, SDS–PAGE/Coomassie blue stain of the purified complex concentrated from the Superdex 200 Increase 10/30 column. For gel source data, see Supplementary Fig. 1. **d**, Representative elution profile of Flag-purified complex on Superdex 200 Increase 10/30 SEC.

## Extended Data Fig. 4 Cryo-EM data processing for the MIPS521–ADO $\beta$ –A<sub>1</sub>R–G<sub>i2</sub> and ADO–A<sub>1</sub>R–G<sub>i2</sub> complexes.

**a**, MIPS521–ADO $\beta$ –A<sub>1</sub>R–G<sub>i2</sub>; **b**, ADO–A<sub>1</sub>R–G<sub>i2</sub>. Representative cryo-EM micrographs of each of the complexes. Reference-free 2D class averages of the complexes in LMNG and CHS detergent micelles. Gold-standard Fourier shell correlation (FSC) curves, showing the overall nominal resolution of 3.2 Å and 3.3 Å, respectively, at FSC 0.143. Corresponding 3D cryo-EM maps coloured according to local resolution estimation (Å) in Relion. **c**, Atomic resolution model of representative regions from the MIPS521–ADO–A<sub>1</sub>R–G<sub>i2</sub> structure of the A<sub>1</sub>R transmembrane domain, ADO, and MIPS521. The molecular model is shown in ball and stick representation, coloured by heteroatom, and the cryo-EM map displayed in mesh contoured at 0.02.

## Extended Data Fig. 5 Stable hydrogen bonds formed between residue S6.47/L7.41 in A<sub>1</sub>R and MIPS521 in A<sub>1</sub>R–G<sub>i2</sub>–MIPS521.

**a, c**, GaMD and **b, d**, cMD simulations. Each simulation trace is displayed in a different colour (black, red, blue). The lines depict the running average over 2 ns.

## Extended Data Fig. 6 Affinity of orthosteric ligands at mutations of the MIPS521 extrahelical allosteric binding pocket.

**a, c**, The affinity of (**a**) [<sup>3</sup>H]-DPCPX and (**c**) NECA for wildtype and mutant A<sub>1</sub>Rs performed in FlpInCHO cells. **b**, Bmax; determined by [<sup>3</sup>H]-DPCPX radioligand saturation binding studies. Data are the means + S.E.M. of 3–7 independent experiments (shown as circles) performed in duplicate. \*P < 0.05 (compared with WT; one-way analysis of variance, Dunnett's post-hoc test).

## Extended Data Fig. 7 Extrahelical binding sites for allosteric modulators of class A GPCRs.

The unique extrahelical binding pose of MIPS521 in the A<sub>1</sub>R (orange) compared to previously reported extrahelical allosteric binding pockets for class A GPCRs in P2Y1R (BPTU, red; PDB 4XNV), PAR2 (AZ3451, yellow; PDB 5NDZ), CB1 (ORG28569, green; 6KQI), GPR40 (AP8, cyan; PDB 5TZY), C5aR (NDT9513727, blue; PDB 5O9H), D1R (LY3154207, navy; PDB 7LJD), and β2AR (Compound-6FA, pink; PDB 6N48).

## Extended Data Fig. 8 Stability of MIPS521 at the allosteric binding site of A<sub>1</sub>R is enhanced by G<sub>i2</sub> protein coupling to the receptor.

**a, b**, RMSD (Å) of MIPS521 relative to the starting cryo-EM conformation obtained from GaMD simulations in the **(a)** absence and **(b)** presence of G<sub>i2</sub>. **c, d**, RMSD (Å) of MIPS521 relative to the starting cryo-EM conformation obtained from cMD simulations in the **(c)** absence and **(d)** presence of G<sub>i2</sub>. Each condition represents three GaMD/cMD simulations, with each simulation trace displayed in a different colour (black, red, blue). Lines depict the running average over 2 ns.

## Extended Data Fig. 9 MIPS521 stabilizes the A<sub>1</sub>R–G<sub>i2</sub> ternary complex.

**a–d**, RMSD (Å) of ADO from cMD simulations completed in the **(a)** absence or **(b)** presence of MIPS521, **(c)** G<sub>i2</sub>, or **(d)** both G<sub>i2</sub> and MIPS521. **e–h**, Distance between the intracellular ends of TM3 and TM6 (measured as the distance in Å between Arg105<sup>3.50</sup> and Glu229<sup>6.30</sup>) in the **(e)** absence or **(f)** presence of MIPS521, **(g)** G<sub>i2</sub>, or **(h)** both G<sub>i2</sub> and MIPS521. Each condition represents three cMD simulations, with each simulation trace displayed in a different colour (black, red, blue). The lines depict the running average over 2 ns. **i, j**, Distance between A<sub>1</sub>R and G<sub>i2</sub> (measured as the distance in Å between the NPxxY motif of A<sub>1</sub>R and the C terminus of

the  $\text{G}\alpha\alpha 5$  helix) from GaMD simulations in the (**i**) absence and (**j**) presence of MIPS521. **k, l**, Distance between  $\text{A}_1\text{R}$  and  $\text{G}_{i2}$  from cMD simulations in the (**k**) absence and (**l**) presence of MIPS521. Each condition represents three GaMD/cMD simulations, with each simulation trace displayed in a different colour (black, red, blue). Thick lines depict the running average over 2 ns. **m–p**, Flexibility change upon removal of PAM and/or  $\text{G}_{i2}$  protein from the ADO-bound  $\text{A}_1\text{R}$  obtained from GaMD simulations. (**m**, RMSFs of the  $\text{A}_1\text{R}-\text{G}_{i2}$ -MIPS521. A colour scale of 0.0 Å (blue) to 5.0 Å (red) was used. **n**, Change in the RMSFs of the  $\text{A}_1\text{R}-\text{G}_{i2}$  when MIPS521 was removed from  $\text{A}_1\text{R}-\text{G}_{i2}$ -MIPS521. **o**, Change in the RMSFs of the  $\text{A}_1\text{R}$  and MIPS521 when the  $\text{G}_{i2}$  was removed from  $\text{A}_1\text{R}-\text{G}_{i2}$ -MIPS521. **p**, Change in the RMSFs of the  $\text{A}_1\text{R}$  when the  $\text{G}_{i2}$  and MIPS521 were removed from  $\text{A}_1\text{R}-\text{G}_{i2}$ -MIPS521 system. A colour scale of -2.0 Å (blue) to 2.0 Å (red) was used for **n**, **o** and **p**.

## Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

## Supplementary information

### Supplementary Information

This file contains Supplementary Tables 1–6 and their accompanying legends.

### Reporting Summary

### Supplementary Figure 1

Original western and SDS-PAGE gels used to generate Extended Data Fig. 3b, c. Dotted boxes indicate the area of gel used.

## Rights and permissions

## [Reprints and Permissions](#)

## About this article

### Cite this article

Draper-Joyce, C.J., Bhola, R., Wang, J. *et al.* Positive allosteric mechanisms of adenosine A<sub>1</sub> receptor-mediated analgesia. *Nature* **597**, 571–576 (2021).  
<https://doi.org/10.1038/s41586-021-03897-2>

- Received: 06 November 2020
- Accepted: 11 August 2021
- Published: 08 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03897-2>

### 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/s41586-021-03897-2>

- Matters Arising
- [Published: 22 September 2021](#)

# A finding of sex similarities rather than differences in COVID-19 outcomes

- [Heather Shattuck-Heidorn](#) ORCID: [orcid.org/0000-0002-8565-4876<sup>1,2</sup>](https://orcid.org/0000-0002-8565-4876),
- [Ann Caroline Danielsen<sup>3</sup>](#),
- [Annika Gompers](#) ORCID: [orcid.org/0000-0002-2591-3477<sup>4</sup>](https://orcid.org/0000-0002-2591-3477),
- [Joseph Dov Bruch<sup>3</sup>](#),
- [Helen Zhao<sup>5</sup>](#),
- [Marion Boulicault<sup>6</sup>](#),
- [Jamie Marsella<sup>7</sup>](#) &
- [Sarah S. Richardson<sup>2,7</sup>](#)

[Nature](#) volume 597, pages E7–E9 (2021)

- 2209 Accesses
- 70 Altmetric
- [Metrics details](#)

## Subjects

- [Immunology](#)
- [Risk factors](#)
- [SARS-CoV-2](#)

Matters Arising to this article was published on 22 September 2021

The Original Article was published on 26 August 2020

Download PDF

arising from T. Takahashi et al. *Nature* <https://doi.org/10.1038/s41586-020-2700-3> (2020)

The sex disparity in COVID-19 mortality varies widely and is of uncertain origin. In their recent Article, Takahashi et al.<sup>1</sup> assess immune phenotype in a sample of patients with COVID-19 and conclude that the “immune landscape in COVID-19 patients is considerably different between the sexes”, warranting different vaccine and therapeutic regimes for men and women—a claim that was disseminated widely following the publication<sup>2</sup>. Here we argue that these inferences are not supported by their findings and that the study does not demonstrate that biological sex explains COVID-19 outcomes among patients. The study overstates its findings and factors beyond innate sex are treated superficially in analysing the causes of gender or sex disparities in COVID-19 disease outcomes.

Takahashi et al. measured more than 100 immune markers in a sample of patients with COVID-19 and uninfected healthcare workers (HCW). They compared male and female patients and HCW both at baseline and longitudinally over the disease course. These comparative analyses, both within sex and between sex, across patients and HCW, at baseline and over time, yielded more than 500 findings<sup>1</sup>. Most of the findings in the paper are presented as raw data, unadjusted for possible covariates. Among the more than 200 findings from adjusted analyses, 13 (6%) remained statistically significant after controlling for covariates (primarily age and body mass index (BMI)). This count excludes analyses on antibodies and viral load, as well as comparisons of female HCW (F\_HCW) versus male HCW (M\_HCW), female patients (F\_Pt) versus female HCW and male patients (M\_Pt) versus male HCW.

There is considerable mismatch between the claims made in the paper and the results presented in the data tables, making it challenging to understand the basis of many of these claims. The discussion section focuses on claims related to ten immune markers, positing a variety of sex differences across diverse analyses (reconstructed in Table 1). The expanded data tables demonstrate that nine of these claims are based on raw data and do not hold true in adjusted analyses. For example, interleukin-18 (IL-18) and IL-8, emphasized in the abstract and discussion as higher in male patients, show a sex difference only in baseline-unadjusted analyses of the smaller cohort. This indicates that these reported sex differences in immunological response are better explained by factors other than biological sex.

### **Table 1 Sex difference claims in Takahashi et al.**

Similarly, attempting to address the potential role of these markers in disparate outcomes between men and women, Takahashi et al. associate lower levels of activated T cells at baseline with poorer outcomes among men, but not among women, in a subsample of 12 patients who deteriorated during the course of the disease (6 male and 6 female). However, as fig. 4 demonstrates, deteriorated male patients are older<sup>1</sup>. After adjusting for age, there are no sex differences in activated T cells among the patient samples.

Although statistical significance is not the only consideration when evaluating study results, the authors use statistical significance to summarize their own results and imply that the central findings remain statistically significant after adjustment. Particularly considering the sweeping scope of the study's conclusions, combined with the study's limited sample size, large confidence intervals, few repeat measures for many participants in the longitudinal cohort, and lack of clinical discussion of effect sizes, statistical significance remains an important guidepost for contextualizing the study's findings.

Three findings that are described as sex differences<sup>1</sup> are actually differences within sexes that do not correspond with between-sex differences (Table 1). For example, CCL5 differs at baseline between female patients who would later deteriorate (F\_deteriorated) and those who remained stable (F\_stabilized) ( $n = 5$  F\_deteriorated; 14 F\_stabilized, adjusted difference:

0.39, 95% confidence interval (0.03, 0.74),  $P = 0.03$ ), with no such difference among male patients who deteriorated and those who remained stable ( $n = 6$  M\_deteriorated; 10 M\_stabilized, adjusted difference: 0.16, 95% confidence interval (-0.23, 0.54),  $P = 0.70$ ). However, comparing the difference-in-difference, there is no evidence that the change in CCL5 between deteriorated and stabilized patients differs between the sexes (adjusted difference: 0.23, 95% confidence interval (-0.18, 0.64),  $P = 0.25$ ). Such within-sex differences without accompanying between-sex differences cannot be interpreted as indicating sex-specific disease progression between men and women.

Overall, Takahashi et al. present three findings that are significant after adjustment and can properly be conceptualized as sex differences<sup>1</sup>: at baseline, numbers of non-classical monocytes (ncMono) were higher in male patients ( $n = 21$  female and 16 male) and activated CD8 T cell numbers were higher in female patients ( $n = 21$  female and 16 male), and male patients had higher levels of CCL5 in longitudinal analysis ( $n = 48$  female and 43 male) (Table 1).

There are also three findings of a greater difference-in-difference that maintain significance after adjustment: at baseline, IL-8 was higher in both male and female patients compared with HCW, but the increase in IL-8 in male patients relative to male HCW was greater than the increase in female patients relative to female HCW ( $n = 19$  F\_Pt, 28 F\_HCW, 16 M\_Pt and 15 M\_HCW); at baseline, CXCL-10 was higher in both male and female patients compared to HCW, but the increase in male patients relative to male HCW was greater than the increase in female patients relative to female HCW ( $n = 19$  F\_Pt, 28 F\_HCW, 16 M\_Pt and 15 M\_HCW); and, in longitudinal analyses, CCL5 increased in male patients compared with male HCW, but did not differ between female patients and female HCW ( $n = 48$  F\_Pt, 28 F\_HCW, 43 M\_Pt and 15 M\_HCW) (Table 1).

However, none of these findings of sex differences appear robust across the conducted analyses. For instance, while baseline levels of ncMono and CD8 T cells differ in the direct comparison between female and male patients, the sex difference disappears in the corresponding difference-in-differences analysis. In addition, none of the markers that do show sex differences in

cohorts A and B emerge as predictive variables of interest in analyses comparing stable with deteriorated patients. While we fully recognize that immune differences would not necessarily be expected to be consistent across analyses, the lack of consistency, illustrated in Table 1, is part of a triangulating web of observations suggesting that the sex difference findings do not show a strong signal and may be artefactual.

Biological sex differences are the only causal model considered in the study. While it is plausible that sex-related biological variables may have a role in explaining sex disparities in COVID-19, strong evidence not cited by the researchers suggests a large role for social and other variables in producing the sex differences they seek to explain. For example, research demonstrates substantial variation in the magnitude and direction of the COVID-19 sex disparity across geographical localities, amongst racial and ethnic groups, and over time; these patterns are better explained by contextual factors than biological sex differences<sup>3,4,5,6</sup>. Previous research also predicts that occupational sex segregation<sup>7</sup> and comorbidities are likely to largely explain COVID-19 sex disparities, as observed in recent SARS-CoV-1 and Middle East respiratory syndrome (MERS) epidemics<sup>8,9,10</sup>. Other studies document gender differences in conformity to COVID-19 public health guidelines<sup>11</sup>. Further research raises questions about whether aggregate patterns of higher COVID-19 mortality in men constitute a COVID-19-specific sex disparity, given men's pre-existing higher aggregate mortality rates before the pandemic<sup>12</sup>.

Gender influences both exposure to the virus and susceptibility to severe outcomes. Occupational work segregation or adherence to behaviours such as mask wearing mediate viral load and therefore disease severity<sup>13</sup>. Chronic diseases, which are differentially distributed across men and women due to both gender- and sex-related factors, are also important contributors to COVID-19 progression and outcomes<sup>14</sup>. Notably, immune function is modified during the progression of many chronic diseases<sup>15</sup>. This is one avenue by which observed differences in immune markers may reflect gendered chronic conditions and associated immune responses rather than sex-specific biological mechanisms in response to the SARS-CoV-2 virus.

In these ways, the claims<sup>1</sup> that sex differences in immune factors underlie COVID-19 sex disparities and merit “sex-dependent approaches to prognosis, prevention, care, and therapy for patients with COVID-19” are not only unsupported by the data, they are also not appropriately contextualized within the empirical literature on the primary role of social factors as causes of sex disparities in respiratory infectious disease epidemics.

The study by Takahashi et al.<sup>1</sup> should be characterized as an exploratory study of possible associations between immunological variables and sex disparities in COVID-19 outcomes. The study presents largely null findings that support an assessment of male–female similarities in immune response to the SARS-CoV-2 virus. We stress that in no way does this study provide a foundation for clinical practice or for public health strategies to ameliorate COVID-19 sex disparities.

## References

1. 1.

Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).

2. 2.

Mandavilli, A. Why does the coronavirus hit men harder? A new clue. *The New York Times*  
<https://www.nytimes.com/2020/08/26/health/coronavirus-men-immune.html> (26 August 2020).

3. 3.

US gender/sex COVID-19 data tracker. *Gender Sci Lab*  
<https://www.genderscilib.org/gender-and-sex-in-covid19> (2020).

4. 4.

Islam, N., Khunti, K., Dambha-Miller, H., Kawachi, I. & Marmot, M. COVID-19 mortality: a complex interplay of sex, gender and ethnicity. *Eur. J. Public Health* **30**, 847–848 (2020).

5. 5.

Dehingia, N. & Raj, A. Sex differences in COVID-19 case fatality: do we know enough? *Lancet* **9**, e14–e15 (2021).

6. 6.

Rushovich, T. et al. Sex disparities in COVID-19 mortality vary across US racial groups. *J. Gen. Intern. Med.* **36**, 1696–1701 (2021).

7. 7.

Adams, R. B. Gender equality in work and COVID-19 deaths. *Covid Economics* (11 May 2020).

8. 8.

Chan-Yeung, M. & Xu, R. H. SARS: epidemiology. *Respirology* **8** (Suppl), S9–S14 (2003).

9. 9.

Jia, N. et al. Case fatality of SARS in mainland China and associated risk factors. *Trop. Med. Int. Health* **14** (Suppl 1), 21–27 (2009).

10. 10.

Yang, Y.-M. et al. Impact of comorbidity on fatality rate of patients with Middle East respiratory syndrome. *Sci. Rep.* **7**, 11307–11309 (2017).

11. 11.

Galasso, V., et al. *Gender Difference in COVID-19 Related Attitudes and Behavior: Evidence from a Panel Survey in Eight OECD*

*Countries*; <https://www.nber.org/papers/w27359> (National Bureau of Economic Research, 2020).

12. 12.

Krieger, N., Chen, J. T. & Waterman, P. D. Excess mortality in men and women in Massachusetts during the COVID-19 pandemic. *Lancet* **395**, 1829 (2020).

13. 13.

Gandhi, M., Beyrer, C. & Goosby, E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. *J. Gen. Intern. Med.* **35**, 3063–3066 (2020).

14. 14.

Zhang, J. et al. Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. *Clin. Microbiol. Infect.* **26**, 767–772 (2020).

15. 15.

Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).

## Author information

### Affiliations

1. Women and Gender Studies Program, University of Southern Maine, Portland, ME, USA

Heather Shattuck-Heidorn

2. Studies of Women, Gender and Sexuality, Harvard University, Cambridge, MA, USA

Heather Shattuck-Heidorn & Sarah S. Richardson

3. Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Ann Caroline Danielsen & Joseph Dov Bruch

4. Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, MA, USA

Annika Gompers

5. Department of Philosophy, Columbia University, New York, NY, USA

Helen Zhao

6. Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA

Marion Boulicault

7. Department of the History of Science, Cambridge, MA, USA

Jamie Marsella & Sarah S. Richardson

## **Contributions**

All authors contributed to conceptualization and design. H.S.H., A.G., A.C.D., J.D.B. and S.S.R. performed the analysis. H.S.H. and S.R. drafted the manuscript, and all authors contributed to revisions and editing.

## **Corresponding author**

Correspondence to Heather Shattuck-Heidorn.

## **Ethics declarations**

## Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Shattuck-Heidorn, H., Danielsen, A.C., Gompers, A. *et al.* A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* **597**, E7–E9 (2021). <https://doi.org/10.1038/s41586-021-03644-7>

- Received: 13 November 2020
- Accepted: 11 May 2021
- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03644-7>

## 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/s41586-021-03644-7>

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

- Matters Arising
- [Published: 22 September 2021](#)

# Reply to: A finding of sex similarities rather than differences in COVID-19 outcomes

- [Takehiro Takahashi](#)<sup>1</sup>,
- [Mallory K. Ellingson](#)<sup>2</sup>,
- [Patrick Wong](#)<sup>1</sup>,
- [Benjamin Israelow](#)<sup>1,3</sup>,
- [Carolina Lucas](#)<sup>1</sup>,
- [Jon Klein](#)<sup>1</sup>,
- [Julio Silva](#)<sup>1</sup>,
- [Tianyang Mao](#)<sup>1</sup>,
- [Ji Eun Oh](#)<sup>1</sup>,
- [Maria Tokuyama](#)<sup>1</sup>,
- [Peiwen Lu](#)<sup>1</sup>,
- [Aryind Venkataraman](#)<sup>1</sup>,
- [Annsea Park](#)<sup>1</sup>,
- [Feimei Liu](#)<sup>1,4</sup>,
- [Amit Meir](#)<sup>5</sup>,
- [Jonathan Sun](#)<sup>6</sup>,
- [Eric Y. Wang](#)<sup>1</sup>,
- [Arnau Casanovas-Massana](#)<sup>2</sup>,
- [Anne L. Wyllie](#)<sup>2</sup>,
- [Chantal B. F. Vogels](#)<sup>2</sup>,
- [Rebecca Ernest](#)<sup>2</sup>,
- [Sarah Lapidus](#)<sup>2</sup>,
- [Isabel M. Ott](#)<sup>2,7</sup>,

- [Adam J. Moore](#)<sup>2</sup>,
- [Albert Shaw](#)<sup>3</sup>,
- [John B. Fournier](#)<sup>3</sup>,
- [Camila D. Odio](#)<sup>3</sup>,
- [Shelli Farhadian](#)<sup>3</sup>,
- [Charles Dela Cruz](#)<sup>8</sup>,
- [Nathan D. Grubaugh](#)<sup>2</sup>,
- [Wade L. Schulz](#)<sup>9,10</sup>,
- [Aaron M. Ring](#)<sup>1</sup>,
- [Albert I. Ko](#)<sup>2</sup>,
- [Saad B. Omer](#) <sup>2,3,11,12</sup> &
- [Akiko Iwasaki](#) <sup>1,13</sup>

*Nature* volume **597**, pages E10–E11 (2021)

- 1461 Accesses
- 5 Altmetric
- [Metrics details](#)

## Subjects

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

The [Original Article](#) was published on 22 September 2021

[Download PDF](#)

replying to H. Shattuck-Heidorn et al. *Nature*  
<https://doi.org/10.1038/s41586-021-03644> (2021)

In the accompanying Comment, Shattuck-Heidorn et al.<sup>1</sup> argue that in our study<sup>2</sup> the inferences are not supported by the data and the study is not appropriately contextualized within the empirical literature on the primary role of social factors in infectious disease epidemics. Our study should be read in the context of the large body of studies on the biological sex differences of immune responses. Many studies have shown that human immune responses against infections differ between the sexes<sup>3</sup>, and this is also the case in COVID-19<sup>4,5</sup>. Such evidence in human studies is supported by a large body of animal studies that are devoid of any confounding social, behavioural and demographic factors, demonstrating that there are sex differences in immune responses across species, from fruitflies to mice<sup>3</sup>. In a mouse model of SARS-CoV, female mice are protected owing to the influence of female sex hormones on the immune system<sup>6</sup>. A recent study using a mouse model of SARS-CoV-2 infection also demonstrated a significant survival advantage in female mice<sup>7</sup>; male mice produce larger inflammatory responses with significantly higher expression of gene signatures of crucial cytokines and chemokines compared with female mice<sup>7</sup>, which is in line with our findings<sup>2</sup>. The role of sex and gender in the causal pathway is complex along the time course of infection (exposure, symptomatic illness, moderate and severe disease), and it involves biological and contextual factors. However, the purpose of our study was to examine the role of biological sex in immune responses among hospitalized patients, for which there is evidence of significant gender-based differences<sup>8</sup>.

Nevertheless, we are mindful of the limitations of our study, such as the small sample sizes, and of its exploratory nature; however, we disagree with the conclusion<sup>1</sup> that our study “presents largely null findings that support an assessment of male–female similarities in immune response to the SARS-CoV-2 virus”.

Shattuck-Heidorn et al. constructed table 1<sup>1</sup> from our extended data tables 3–6<sup>2</sup> by classifying the data according to whether it was significant (that is,  $P < 0.05$ ). They argue that some of the data that are significant in the baseline analysis are no longer significant after adjusting for age and body mass index (BMI), and that the factors in which there were statistically

significant differences in the baseline analysis and the longitudinal analysis are not the same, suggesting a lack of consistency. Furthermore, the authors argue that differences reported in our study are largely null and maybe even artefactual.

Our study was an exploratory, and not a hypothesis-driven analysis, with a small sample size to provide a basis for further investigations. Therefore, although we used significance testing in our own interpretations, it is wrong to interpret any results that are not statistically significant results as disproving a hypothesis<sup>1</sup>—that is, to suggest that a lack of statistical significance indicates that there is no effect. *P* values are a useful tool but, as has been thoroughly discussed in the biostatistical literature<sup>9</sup>, it is inappropriate to interpret them in isolation from effect sizes, sample size and study design. Arguments based solely on *P* values lead to the dismissal of important differences. For example, by evaluating the magnitude and direction of the unadjusted and adjusted differences, as well as the statistical significance, in IL-8 and IL-18 levels between male and female patients in cohort A, we identified an important difference, which has been confirmed by others as discussed below. In addition, Shattuck-Heidorn et al.<sup>1</sup> argue that significant differences in numbers of activated CD8 T cells between stable and deteriorated males disappears after adjusting for age and BMI. This is exactly what is expected—we clearly showed that deteriorated males were older, and exhibited lower T cell activation, and that these factors were strongly correlated only in males.

The claim that the factors in which statistical significance is detected in the baseline and longitudinal analyses should be ‘consistent’<sup>1</sup> is based on an assumption that the same immune factors should be found in different phases of COVID-19 infection. Baseline analysis of cohort A included only the first time point, only for patients with moderate disease. The longitudinal analysis of cohort B included samples from later disease phases, with varying severity, and takes into account the overall immunological changes throughout the course of the disease. The immune response is a dynamic process involving innate and adaptive immunity<sup>10</sup>, and cytokine levels may change by orders of magnitude over time<sup>11</sup>. Thus, these analyses are fundamentally asking different questions and would not be expected to identify the same factors.

We are confused by the authors' claim that the differences in immune phenotypes are largely null on the basis of biological sex, while at the same time they state that "observed differences in immune markers may reflect gendered chronic conditions". The biological sex differences are closely intertwined with differences due to social and demographic gender disparities, and they are not mutually exclusive. We agree that analyses of the impact of gender disparities on immune responses are very important. However, this was not the focus of our study. We explicitly focused on the biological sex differences in the COVID-19 immune responses among a defined set of patients, and did not make general claims about the biological bases of gender disparities.

In less than half a year since the publication of our study, a large body of literature is emerging to support our findings. A single-cell transcriptomic study of peripheral blood mononuclear cells from patients with COVID-19 has revealed a significantly higher abundance of non-classical monocytes (ncMono) in male patients compared with female patients<sup>12</sup>, as we reported in our baseline analysis<sup>2</sup>, which is being dismissed by Shattuck-Heidorn et al. in their table 1<sup>1</sup>. The ncMono abundance in male patients was twofold to fourfold higher compared with female patients<sup>12</sup>—the same magnitude of difference as in our study<sup>2</sup>. In addition, *IL18* expression in monocytes from male patients was significantly higher than in those from female patients<sup>12</sup>. Nasal squamous epithelial cells from male patients with COVID-19 also expressed higher levels of *IL18* than those from female patients<sup>12</sup>. Male patients showed higher expression in monocytes of *MYD88* and *NFKB1*<sup>12</sup>, genes that encode direct regulators of pro-inflammatory cytokines including IL-8. The neutrophil:lymphocyte ratio was found to be higher in male patients<sup>13</sup>, and neutrophil activation was associated with IL-8 levels in patients with COVID-19<sup>14</sup>. Another study used single-cell RNA-sequencing analysis to demonstrate prominent sex differences in CD8 T cells and especially in the subpopulation of CD161<sup>hi</sup> mucosal-associated invariant T cells (MAIT cells)<sup>15</sup>. MAIT cells in males exhibited pro-apoptotic gene signatures, whereas the same cell type in females had a different set of activated gene signatures, and bioinformatic analysis of gene-expression patterns indicated that these cells interact with monocytes through CCL5–CCR1 and IL18–IL-18R ligand–receptor interactions<sup>15</sup>; IL-18 and CCL5

are the same factors for which we reported sex differences in our baseline and longitudinal analyses, respectively<sup>2</sup>. The striking concordance between our findings and others on sex differences that implicate the same parameters and associated immune pathways makes it highly unlikely that our findings are artefactual. Independent studies including ours, using different modalities and methods, support sex differences in the same immune factors and pathways.

Finally, referring to our study<sup>2</sup>, Shattuck-Heidorn et al. state<sup>1</sup> that “We stress that in no way does this study provide a foundation for clinical practice or for public health strategies to ameliorate COVID-19 sex disparities”. We simply stated that our analyses “provide a potential basis for taking sex-dependent approaches to prognosis, prevention, care, and therapy for patient with COVID-19”. Science is an iterative process. Although our study in isolation may only contribute a piece of the puzzle, given the large body of studies that demonstrate sex differences during the course of COVID-19 disease and the immune response as outlined above, it is perhaps time to take these collective insights into account for future guidance in developing clinical practice and public health strategies to improve treatment and prevention for COVID-19.

In conclusion, accumulating evidence supports an important role for biological sex in immune responses against COVID-19. The heterogeneity in the disease phenotype in COVID-19 is related to the intersectional nature of a variety of factors—social, gender, race, ethnicity, disability and economic, as well as geography, age and comorbidities<sup>16</sup>. We believe that biological sex should be included as a key variable for studying infectious diseases. We hope that more studies in this area will contribute to the better understanding of disease mechanisms, as well as to the development of better treatments against acute and long COVID-19.

## References

1. 1.

Shattuck-Heidorn, H. et al. A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature*

<https://doi.org/10.1038/s41586-021-03644-7> (2021).

2. 2.

Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).

3. 3.

Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).

4. 4.

Bunders, M. J. & Altfeld, M. Implications of sex differences in immunity for SARS-CoV-2 pathogenesis and design of therapeutic interventions. *Immunity* **53**, 487–495 (2020).

5. 5.

Klein, S. L. et al. Biological sex impacts COVID-19 outcomes. *PLoS Pathog.* **16**, e1008570 (2020).

6. 6.

Channappanavar, R. et al. Sex-based differences in susceptibility to severe acute respiratory syndrome coronavirus infection. *J. Immunol.* **198**, 4046–4053 (2017).

7. 7.

Golden, J. W. et al. Human angiotensin-converting enzyme 2 transgenic mice infected with SARS-CoV-2 develop severe and fatal respiratory disease. *JCI Insight* **5**, e142032 (2020).

8. 8.

Gebhard C et al. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol. Sex Differ.* **11**, 29 (2020).

9. 9.

Wasserstein, R. et al. The ASA Statement on *P*-values: context, process, and purpose. *Am. Stat.* **70**, 129–133 (2016).

10. 10.

Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat. Immunol.* **16**, 343–353 (2015).

11. 11.

Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).

12. 12.

How, Y. et al. Multimodal single-cell omics analysis of COVID-19 sex differences in human immune systems. Preprint at <https://doi.org/10.1101/2020.12.01.407007> (2020).

13. 13.

Meng Y. et al. Sex-specific clinical characteristics and prognosis of coronavirus disease-19 infection in Wuhan, China: a retrospective study of 168 severe patients. *PLoS Pathog.* **16**, e1008520 (2020).

14. 14.

Meizlish, M. L. et al. A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv.* **5**, 1164–1177 (2021).

15. 15.

Yu, C. et al. Mucosal-associated invariant T cell responses differ by sex in COVID-19. *Med* **2**, 755–772.e5 (2021).

16. 16.

Shapiro, J. R., Klein, S. L. & Morgan, R. COVID-19: use intersectional analyses to close gaps in outcomes and vaccination. *Nature* **591**, 202 (2021).

## Author information

### Affiliations

1. Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA

Takehiro Takahashi, Patrick Wong, Benjamin Israelow, Carolina Lucas, Jon Klein, Julio Silva, Tianyang Mao, Ji Eun Oh, Maria Tokuyama, Peiwen Lu, Arvind Venkataraman, Annsea Park, Feimei Liu, Eric Y. Wang, Aaron M. Ring & Akiko Iwasaki

2. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA

Mallory K. Ellingson, Arnau Casanovas-Massana, Anne L. Wyllie, Chantal B. F. Vogels, Rebecca Earnest, Sarah Lapidus, Isabel M. Ott, Adam J. Moore, Nathan D. Grubaugh, Albert I. Ko & Saad B. Omer

3. Department of Medicine, Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT, USA

Benjamin Israelow, Albert Shaw, John B. Fournier, Camila D. Odio, Shelli Farhadian & Saad B. Omer

4. Department of Biomedical Engineering, Yale School of Engineering and Applied Science, New Haven, CT, USA

Feimei Liu

5. Boyer Center for Molecular Medicine, Department of Microbial Pathogenesis, Yale University, New Haven, CT, USA

Amit Meir

6. Department of Comparative Medicine, Yale University School of Medicine, New Haven, CT, USA

Jonathan Sun

7. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

Isabel M. Ott

8. Department of Medicine, Section of Pulmonary and Critical Care Medicine, Yale University School of Medicine, New Haven, CT, USA

Charles Dela Cruz

9. Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT, USA

Wade L. Schulz

10. Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA

Wade L. Schulz

11. Yale Institute for Global Health, Yale University, New Haven, CT, USA

Saad B. Omer

12. Yale School of Nursing, Yale University, Orange, CT, USA

Saad B. Omer

13. Howard Hughes Medical Institute, Chevy Chase, MD, USA

Akiko Iwasaki

## Contributions

T.T., M.K.E., S.B.O., and A.I. drafted the manuscript. All authors helped to edit the manuscript. The Yale IMPACT Research Team authors included in the original Article contributed to sample collection, processing, raw data acquisition, and the creation and design of Yale COVID-19 study cohort, but not to the analysis of the data, so these authors have not been included in this Reply.

## Corresponding authors

Correspondence to Takehiro Takahashi or Mallory K. Ellingson or Saad B. Omer or Akiko Iwasaki.

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Rights and permissions

### [Reprints and Permissions](#)

## About this article

### Cite this article

Takahashi, T., Ellingson, M.K., Wong, P. *et al.* Reply to: A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* **597**,

E10–E11 (2021). <https://doi.org/10.1038/s41586-021-03645-6>

- Published: 22 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03645-6>

## 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

## [\*\*Sex differences in immune responses that underlie COVID-19 disease outcomes\*\*](#)

- Takehiro Takahashi
- Mallory K. Ellingson
- Akiko Iwasaki

Article 26 Aug 2020

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03645-6>

# Amendments & Corrections

- **[Author Correction: CO<sub>2</sub> doping of organic interlayers for perovskite solar cells](#)** [ 03 September 2021]

Author Correction •

- Author Correction
- [Published: 03 September 2021](#)

# Author Correction: CO<sub>2</sub> doping of organic interlayers for perovskite solar cells

- [Jaemin Kong](#) [ORCID: orcid.org/0000-0002-3236-1658](#)<sup>1</sup>,
- [Yongwoo Shin](#) [ORCID: orcid.org/0000-0001-7760-9883](#)<sup>2</sup>,
- [Jason A. Röhr](#) [ORCID: orcid.org/0000-0002-8790-340X](#)<sup>1</sup>,
- [Hang Wang](#)<sup>1</sup>,
- [Juan Meng](#)<sup>1</sup>,
- [Yueshen Wu](#)<sup>3</sup>,
- [Adlai Katzenberg](#) [ORCID: orcid.org/0000-0002-4793-5853](#)<sup>1</sup>,
- [Geunjin Kim](#)<sup>4</sup>,
- [Dong Young Kim](#)<sup>5</sup>,
- [Tai-De Li](#)<sup>6,7</sup>,
- [Edward Chau](#) [ORCID: orcid.org/0000-0001-6970-9390](#)<sup>1</sup>,
- [Francisco Antonio](#)<sup>8</sup>,
- [Tana Siboonruang](#)<sup>1</sup>,
- [Sooncheol Kwon](#)<sup>9</sup>,
- [Kwanghee Lee](#)<sup>10,11</sup>,
- [Jin Ryoun Kim](#) [ORCID: orcid.org/0000-0001-6156-8730](#)<sup>1</sup>,
- [Miguel A. Modestino](#) [ORCID: orcid.org/0000-0003-2100-7335](#)<sup>1</sup>,
- [Hailiang Wang](#) [ORCID: orcid.org/0000-0003-4409-2034](#)<sup>3</sup> &
- [André D. Taylor](#) [ORCID: orcid.org/0000-0003-3241-8543](#)<sup>1,8</sup>

[Nature](#) volume 597, page E12 (2021)

- 1122 Accesses

- 5 Altmetric
- [Metrics details](#)

## Subjects

- [Photocatalysis](#)
- [Solar cells](#)

The [Original Article](#) was published on 02 June 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03518-y>  
Published online 2 June 2021

In this Article, André D. Taylor's affiliation was mistakenly shown with a second affiliation to City University of New York. His correct affiliations are with the Department of Chemical and Biomolecular Engineering, New York University Tandon School of Engineering, New York, NY, USA and the Department of Chemical and Environmental Engineering, Yale University, New Haven, CT, USA. The original Article has been corrected online.

## Author information

### Affiliations

1. Department of Chemical and Biomolecular Engineering, New York University Tandon School of Engineering, New York, NY, USA

Jaemin Kong, Jason A. Röhr, Hang Wang, Juan Meng, Adlai Katzenberg, Edward Chau, Tana Siboonruang, Jin Ryoun Kim, Miguel A. Modestino & André D. Taylor

2. Advanced Materials Laboratory, Samsung Semiconductor, Inc,  
Cambridge, MA, USA

Yongwoo Shin

3. Department of Chemistry, Yale University, New Haven, CT, USA

Yueshen Wu & Hailiang Wang

4. Division of Advanced Materials, Korea Research Institute of Chemical  
Technology (KRICT), Daejeon, Republic of Korea

Geunjin Kim

5. Samsung Advanced Institute of Technology, Samsung Electronics,  
Suwon, Republic of Korea

Dong Young Kim

6. Advanced Science Research Center, The Graduate Center of the City  
University of New York, New York, NY, USA

Tai-De Li

7. Department of Physics, City College of New York, New York, NY,  
USA

Tai-De Li

8. Department of Chemical and Environmental Engineering, Yale  
University, New Haven, CT, USA

Francisco Antonio & André D. Taylor

9. Department of Carbon Convergence Engineering, Wonkwang  
University, Iksan, Republic of Korea

Sooncheol Kwon

10. Heeger Center for Advanced Materials (HCAM), Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea

Kwanghee Lee

11. Research Institute for Solar and Sustainable Energies (RISE), Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea

Kwanghee Lee

## Corresponding author

Correspondence to André D. Taylor.

## Rights and permissions

[Reprints and Permissions](#)

## About this article

### Cite this article

Kong, J., Shin, Y., Röhr, J.A. *et al.* Author Correction: CO<sub>2</sub> doping of organic interlayers for perovskite solar cells. *Nature* **597**, E12 (2021). <https://doi.org/10.1038/s41586-021-03839-y>

- Published: 03 September 2021
- Issue Date: 23 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03839-y>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03839-y>.

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