

nature

AGENTS OF DECAY

How insects aid the decomposition of forest deadwood



After the ban

What lies beyond two weeks for lab-grown human embryos?

Nutritional benefit

Science must help fix global food systems to end hunger

Chiral nitrogen

Stereoselective synthesis of ammonium ions

Nature.2021.09.04

[Sat, 04 Sep 2021]

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

This Week

- **[The world's scientific panel on biodiversity needs a bigger role](#)** [31 August 2021]
Editorial • IPBES, the international panel of leading biodiversity researchers, should be consulted on how best to measure species loss.
- **[The global research community must not abandon Afghanistan](#)** [01 September 2021]
Editorial • Here's how Afghanistan's scholars can be supported.
- **[Demand five precepts to aid social-media watchdogs](#)** [31 August 2021]
World View • Facebook's heavy-handedness could open a window for research in the public interest.
- **[Electric-blue phosphorescence made easy](#)** [23 August 2021]
Research Highlight • Crystal engineering produces sustained molecular glows at room temperature.
- **[So much ice is melting that Earth's crust is moving](#)** [24 August 2021]
Research Highlight • As the continents' frozen burden dissipates, the ground deforms — not only in the immediate area, but also in far-flung locations.
- **[Ozone hole's healing triggers winds of change](#)** [23 August 2021]
Research Highlight • Shifts in polar weather have accompanied the replenishment of the ozone layer.
- **[A plundered pterosaur reveals the extinct flyer's extreme headgear](#)** [25 August 2021]
Research Highlight • A seized fossil reveals the full glory of a winged reptile from the age of the dinosaurs.
- **[Giant clams have a growth spurt — thanks to pollution](#)** [24 August 2021]
Research Highlight • Human-made particles help the huge molluscs to achieve their massive size faster than their ancestors did.

- **Scattered light yields full picture of tiny motions** [25 August 2021]
Research Highlight • Optical technique overcomes limitations of standard imaging methods.
- **Why gassy planets are bigger around more-massive stars**
[25 August 2021]
Research Highlight • During formation, the planets can draw a greater amount of hydrogen and helium around themselves, causing them to increase in size.

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

- EDITORIAL
- 31 August 2021

The world's scientific panel on biodiversity needs a bigger role

IPBES, the international panel of leading biodiversity researchers, should be consulted on how best to measure species loss.



A baby green sea turtle in Madagascar, one of the regions where the probability of widespread biodiversity loss is greatest. Credit: Alexis Rosenfeld/Getty

For more than 30 years, the international community has tried and failed to find a path to slow down — and eventually reverse — worldwide declines in the richness of plant and animal species. Next year, it will have another

chance. The 15th Conference of the Parties (COP 15) to the United Nations Convention on Biological Diversity, recently delayed for the third time, is now slated to take place in person in Kunming, China, in April and May 2022.

Biodiversity is fundamental to Earth's life-support systems, and humans depend on the services that nature provides. In 2010, countries committed to slowing the overall rate of biodiversity loss by 2020. But just 6 out of the 20 targets that were agreed on that occasion — at COP 10 in Aichi, Japan — have been even partially met, notable among them a commitment to conserve 17% of the world's land and inland waters.

Ahead of the Kunming meeting, policymakers and scientists are discussing a new action plan, called the Global Biodiversity Framework, which they hope to agree next year. The latest draft (published in July; see go.nature.com/3kbvspd) includes a promise to conserve 30% of the world's land and sea areas by 2030 and reiterates the need to meet earlier targets, including the provision of greater financial support to low-income countries to help them to protect their biodiversity.

Missing link

Researchers around the world are advising on the plan, through the UN's institutions and through universities and various scientific networks. But one piece of the puzzle is missing. In 2012, a host of governments established the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). It periodically reviews the literature and provides summaries of the latest knowledge. However, the countries organizing the COP are not involving IPBES in the action plan in the way that the UN Intergovernmental Panel on Climate Change has been consulted for advice ahead of climate COPs. It is important that IPBES be asked, because policymakers are being presented with a range of ideas that would benefit from the systematic evaluation that a global scientific advisory body would bring.



The world's species are playing musical chairs: how will it end?

For example, biodiversity terminology is often unfamiliar, and therefore challenging, for most policymakers. The word itself — defined by the biodiversity convention as the variety and variability of life on Earth, at the level of genes, species and ecosystems — is not commonly used, nor well understood beyond the scientific community. The magnitude of biodiversity's value to the planet and to people, as well as the risks of losing it, are also not widely appreciated.

Over the years, various teams of scientists have researched and offered ideas on how to communicate the state of biodiversity both accurately and in a way that is accessible and engages the wider public. Some are advocating a biodiversity equivalent of the 1.5°C warming target, or of net-zero emissions. One suggestion, published last year, is for the international community to adopt a target for limiting species extinctions. The goal would be to keep extinctions of known species to below 20 per year globally for the next 100 years — a single headline number to represent biodiversity ([M. D. A. Rounsevell et al. Science 368, 1193–1195; 2020](#)).

A focus on species extinctions as a proxy for biodiversity is not a new idea, and is controversial. However, the authors say that their intention is not to replace biodiversity's many facets with only one number, but to communicate biodiversity in a way that would resonate with more people.

Another group is proposing a composite index — a single score made up of measures of some of biodiversity's main components, including the health of species and ecosystems, as well as the services that biodiversity provides to people, such as pollination and clean water ([C. A. Soto-Navarro et al.](#) [Nature Sustain. https://doi.org/gmjs2f; 2021](#)). This would be biodiversity's equivalent of the UN Human Development Index — first published in 1990 — which amalgamates information on health, education and income into a single number and has been adopted worldwide as a measure of prosperity and well-being.



[Fewer than 20 extinctions a year: does the world need a single target for biodiversity?](#)

A third idea, published by the leaders of some of the world's most influential conservation and environmental science organizations, is called Nature Positive (see [go.nature.com/2ydk89n](#)). Its authors are proposing that the UN's many global environmental agreements should include three common targets: no net loss of nature from 2020 (meaning that although nature might continue to be degraded in some areas, this would be offset by conservation gains elsewhere); some recovery by 2030; and full recovery by 2050. At present, the UN agreements on biodiversity, stopping climate change and combating desertification all have their own processes, occasionally acting together, but more often operating independently. The goal is to get them to sign up to one set of principles.

All of these ideas have advantages and risks, which is why they need to be systematically evaluated by researchers. That's where IPBES's role is crucial. IPBES comprises a broad community of researchers, and, importantly, it represents voices from under-represented low- and middle-income countries, as well as the world's Indigenous peoples. The governments involved in organizing the Kunming COP should ask IPBES to evaluate the ideas being put forward for the next biodiversity action plan, so they can be confident that what they decide has the support of a consensus of researchers, particularly in more-biodiverse regions of the world. Although preparations for the Kunming COP are well under way, this could also happen after the COP.

Biodiversity loss could be as serious for the planet — and for humanity — as climate change. World leaders have become skilled at organizing complex international meetings and making promises that they then fail to keep. The upcoming biodiversity COP risks being one more such event, which is why researchers offering solutions are right to feel frustrated. They should work with IPBES to review their ideas. A unified voice is powerful, and if scientists can present a united front, policymakers will have fewer excuses to continue with business as usual.

Nature **597**, 7–8 (2021)

doi: <https://doi.org/10.1038/d41586-021-02339-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02339-3>

- EDITORIAL
- 01 September 2021

The global research community must not abandon Afghanistan

Here's how Afghanistan's scholars can be supported.



Students from the American University of Afghanistan in Kabul in 2017. The university says it is relocating temporarily to another country. Credit: Wakil Kohsar/AFP/Getty

“The situation in Afghanistan is horrifying. We need immediate assistance.”

This is one of several distressing messages sent to *Nature* by researchers in Afghanistan, following the Taliban’s capture of Kabul on 15 August and the

evacuation of US military forces on 31 August. Researchers are among those who are now especially vulnerable. The United States has been their main source of funding and collaborations, and that puts them at increased risk of persecution by the new rulers. [Most institutions remain closed](#), and many staff and students — women and men — are in hiding.

For now, the Taliban has announced an amnesty, and is urging Afghanistan's professionals to stay in the country and continue to go to work. But researchers interviewed by *Nature* are not taking any risks. Many remember the Taliban's previous rule (1996–2001), and the systematic human-rights violations, particularly against girls, women and minority communities.

Since 2002, Afghanistan has [witnessed a knowledge and information boom](#). New universities have been established, together with some 200 television channels and 1,900 media outlets, according to the United Nations cultural organization UNESCO. The student population has ballooned from 8,000 to 170,000, one-quarter of whom are women. Afghanistan's Academy of Sciences has grown to employ more than 300 people, and its projects include research to create dictionaries in the approximately 40 languages that are spoken in the country. International collaborations have been set up. For example, Kabul University has been working with the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy, to revamp undergraduate physics teaching in Afghanistan.

Since the Taliban's takeover, organizations that help refugee scholars, such as Scholars at Risk, based in New York City, have been calling on universities in various countries to accept faculty members and students who are able to leave. Afghanistan's neighbouring countries in south and central Asia — especially those that have provided higher education for previous generations of Afghanistan's refugee scholars — should also provide support for researchers and students who need it.

But, in a country of 38 million people, most of Afghanistan's researchers will probably be staying. And they, too, need the international research community's support. This will be harder to give. But there are ways in which it can be done, for example by organizing research opportunities in 'neutral' countries — those that are not party to a conflict.

The Jordan-based synchrotron radiation source, SESAME, is an example of one such opportunity. It is designed explicitly to support researchers in countries that have difficult international relations. Although many years in the making, the pay-off has been worth the effort — its participants include Cyprus and Turkey, as well as Iran and Israel. Afghanistan's researchers should now be invited.

During the cold war, the International Institute for Applied Systems Analysis, based near Vienna, was established in 1972 as somewhere that scientists from East and West could collaborate on interdisciplinary global challenges in a neutral country. Both the Soviet Union and the United States were among the founding partners.

The situation in Iran shows what could happen under the alternative scenario — enforcing a policy of isolation. During Iran's Islamic revolution in 1979, a US-backed monarchy was overthrown in a cleric-led takeover. Many of Iran's former Western allies cut off all but basic consular links. Over time, Iran's rulers have become increasingly repressive. This has hit academic communities hard: scientists with global links are arrested, imprisoned and seen as a security threat, as *Nature* and other publications have reported.

That, in turn, has fuelled a brain drain. Researchers at Stanford University in California have estimated that 3 million people — around 4% of the population — had left the country by 2019, and more continue to do so. This compares with half a million who left before 1979 (P. Azadi *et al. Working Pap. 9; Migration and Brain Drain from Iran*, Stanford Univ., 2020). The evidence is clear: Afghanistan's new rulers and the outside world would be unwise if they [pressed replay on the Iranian tape](#).

This means that, to continue supporting Afghanistan's researchers, countries will need to maintain some minimal lines of communication with the new rulers. This will not be easy, and it will need the Taliban to honour its pledge that people who receive US or European funding, or who work with international organizations, will not be persecuted.

Researchers at risk must be able to leave and to resume their lives in countries that can provide them with safety and security. But, at the same time, research leaders in Afghanistan's neighbouring countries — and those

farther afield — must work strenuously to support those Afghans who are staying, and who must not be forgotten or neglected.

Nature **597**, 8 (2021)

doi: <https://doi.org/10.1038/d41586-021-02340-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02340-w>

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

- WORLD VIEW
- 31 August 2021

Demand five precepts to aid social-media watchdogs



Facebook's heavy-handedness could open a window for research in the public interest.

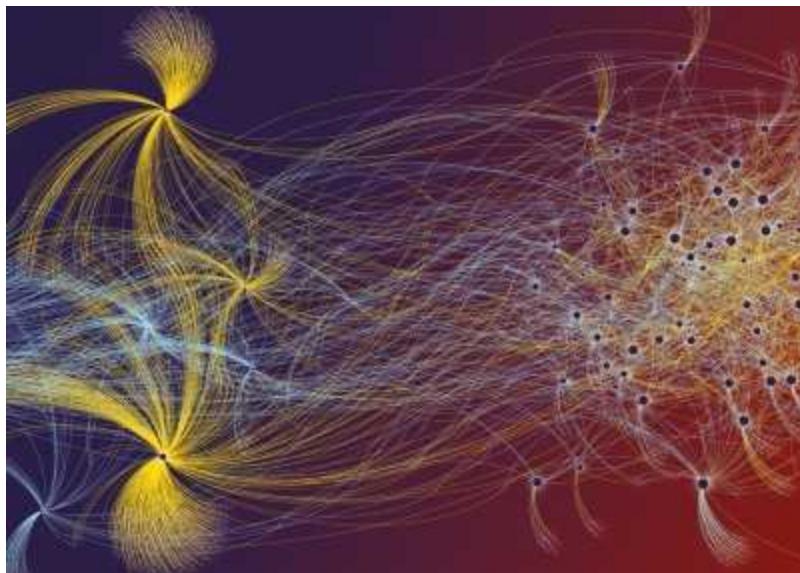
- [Ethan Zuckerman](#) 0

Early this August, Facebook shut down the personal and organizational accounts of researchers associated with New York University's Ad Observatory, a project in which informed volunteers allow study of advertising targeted to their accounts. Facebook said its move was necessary to "protect people's privacy" and to comply with orders from the Federal Trade Commission. The FTC gave an unusually public response. It published a statement saying that its restrictions do not bar "good-faith research in the public interest".

This marks an opportunity for anyone who thinks that social media's effects on democracy and society should be open to scrutiny. It is time to lay down ground rules to empower public-interest research on social media.

In a collaboration with Elizabeth Hansen Shapiro at the Tow Center for Digital Journalism in New York City, I and other colleagues interviewed dozens of researchers, journalists and activists who study how social-media platforms affect democratic participation. Almost all named barriers to data access as a major obstacle, even those who helped to design Social Science One, a highly touted academia–industry partnership to [study the spread of misinformation](#).

Researchers have techniques for dealing with the lack of information the platforms provide, although many such techniques are vulnerable to legal threats or restrictions. Ad Observatory asks for ‘data donation’ from a panel of web users who install a plug-in that allows researchers to study some aspects of the web users’ online activity.



[How Facebook, Twitter and other data troves are revolutionizing social science](#)

Another technique involves scraping — automated collection of content that appears to the general public or logged-in social-media users. This produces data sets such as PushShift, the most comprehensive archive of

content available on the Reddit online discussion forum. Another is Media Cloud, a project I maintain with colleagues at several institutions to index millions of news stories a day and allow study of word frequencies over time. Its automated retrieval and data-storage features are technically identical to a search engine's, and thus prohibited by the non-negotiable terms of service required by most social-media platforms.

Until 2020, the United States' troublingly vague Computer Fraud and Abuse Act made researchers who violated a website's terms of service vulnerable to felony charges. That year, academic researchers argued successfully that using multiple social-media accounts to audit for discrimination should not be considered a criminal activity. A federal court agreed that "mere terms-of-service violations" do not merit criminal charges.

Although the ruling is welcome, uncertainty for researchers remains, and social-media companies actively hinder their work. The FTC's endorsement of 'good-faith research' should be codified into principles guaranteeing researchers access to data under certain conditions.

I propose the following. First, give researchers access to the same targeting tools that platforms offer to advertisers and commercial partners. Second, for publicly viewable content, allow researchers to combine and share data sets by supplying keys to application programming interfaces. Third, explicitly allow users to donate data about their online behaviour for research, and make code used for such studies publicly reviewable for security flaws. Fourth, create safe-haven protections that recognize the public interest. Fifth, mandate regular audits of algorithms that moderate content and serve ads.



[Facebook needs to share more with researchers](#)

In the United States, the FTC could demand this access on behalf of consumers: it has broad powers to compel the release of data. In Europe, making such demands should be even more straightforward. The European Data Governance Act, proposed in November 2020, advances the concept of “data altruism” that allows users to donate their data, and the broader Digital Services Act includes a potential framework to implement protections for research in the public interest.

Technology companies argue that they must restrict data access because of the potential for harm, which also conveniently insulates them from criticism and scrutiny. They cite misuse of data, such as in the Cambridge Analytica scandal (which came to light in 2018 and prompted the FTC orders), in which an academic researcher took data from tens of millions of Facebook users collected through online ‘personality tests’ and gave it to a UK political consultancy that worked on behalf of Donald Trump and the Brexit campaign. Another example of abuse of data is the case of Clearview AI, which used scraping to produce a huge photographic database to allow federal and state law-enforcement agencies to identify individuals.

These incidents have led tech companies to design systems to prevent misuse — but such systems also prevent research necessary for oversight

and scrutiny. To ensure that platforms act fairly and benefit society, there must be ways to protect user data and allow independent oversight.

Part of the solution is to create legal systems, not just technical ones, that distinguish between bad intent and legitimate, public-spirited research that can help to uncover social media's effects on economies and societies.

The influence of social-media companies is undeniable, and executives such as Facebook co-founder Mark Zuckerberg sincerely believe that their platforms make the world a better place. But they have been unwilling to give researchers the data to demonstrate whether this is so. It is time for society to demand access to those data.

Nature **597**, 9 (2021)

doi: <https://doi.org/10.1038/d41586-021-02341-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02341-9>

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

- RESEARCH HIGHLIGHT
- 23 August 2021

Electric-blue phosphorescence made easy

Crystal engineering produces sustained molecular glows at room temperature.

Three images of a world map inkjet printed with the TSP ink under daylight, UV on and UV off conditions

Images of a world map printed with ink made from the phosphorescent crystal in daylight (left), with ultraviolet light on (middle) and with the ultraviolet light switched off (right). Credit: W. Ye *et al.*/*Nature Mater.*

Molecules held in solitary confinement within solid crystals can generate long-lasting glows in various shades of blue.

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**, 10 (2021)

doi: <https://doi.org/10.1038/d41586-021-02282-3>

References

1. 1.

[Nature Mater. \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02282-3>

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

- RESEARCH HIGHLIGHT
- 24 August 2021

So much ice is melting that Earth's crust is moving

As the continents' frozen burden dissipates, the ground deforms — not only in the immediate area, but also in far-flung locations.



Melting on the Greenland ice sheet is contributing to the sideways slip of Earth's crust. Credit: Martin Zwick/REDA&CO/Universal Images Group/Getty

The loss of melting ice from land masses such as Greenland and Antarctica is causing the planet's crust to warp slightly, even in spots more than 1,000 kilometres from the ice loss.

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**, 10 (2021)

doi: <https://doi.org/10.1038/d41586-021-02285-0>

References

1. 1.

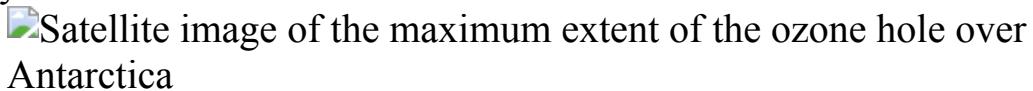
[Geophys. Res. Lett. \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02285-0>

- RESEARCH HIGHLIGHT
- 23 August 2021

Ozone hole's healing triggers winds of change

Shifts in polar weather have accompanied the replenishment of the ozone layer.



As the ozone hole (blue; artist's impression) over the Southern Hemisphere has shrunk, its cooling effect on the upper atmosphere has also dwindled.
Credit: NASA's Goddard Space Flight Centre/SPL

The recovery of the ozone layer over Antarctica has had a palpable effect on temperatures and high-altitude winds around the South Pole.

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**, 10 (2021)

doi: <https://doi.org/10.1038/d41586-021-02284-1>

References

1. 1.

[Nature Geosci. \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02284-1>

- RESEARCH HIGHLIGHT
- 25 August 2021

A plundered pterosaur reveals the extinct flyer's extreme headgear

A seized fossil reveals the full glory of a winged reptile from the age of the dinosaurs.



A newly described fossil of the pterosaur *Tupandactylus navigans* (artist's impression) preserves the soft tissue of its huge head crest, which might have made it clumsier in the air. Credit: Victor Beccari

When police busted a fossil-smuggling operation at Brazil's largest port, they recovered six yellowish limestone slabs — in which the nearly complete remains of an extinct winged reptile called a pterosaur were embedded. What's more, this pterosaur species was formerly known only from skulls.

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**, 10 (2021)

doi: <https://doi.org/10.1038/d41586-021-02283-2>

References

1. 1.

[PLoS ONE \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02283-2>

- RESEARCH HIGHLIGHT
- 24 August 2021

Giant clams have a growth spurt — thanks to pollution

Human-made particles help the huge molluscs to achieve their massive size faster than their ancestors did.



Fluted giant clams (*Tridacna squamosa*) in the Red Sea are growing at the giant clam's equivalent of breakneck speed. Credit: Georgette Douwma/Nature Picture Library/Alamy

Pollution and rising sea temperatures are wiping out delicate coral reef ecosystems. But the conditions that devastate corals haven't ravaged the northern Red Sea's giant clams, which seem to be growing at unprecedented speed.

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**, 11 (2021)

doi: <https://doi.org/10.1038/d41586-021-02281-4>

References

1. 1.

[Proc. R. Soc. B \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02281-4>

- RESEARCH HIGHLIGHT
- 25 August 2021

Scattered light yields full picture of tiny motions

Optical technique overcomes limitations of standard imaging methods.



An innovative imaging technique captures how a gel deforms (arrows and colours) as a crack propagates through it. Credit: S. Aime *et al./Phys. Rev. Lett.*

Researchers have developed an optical imaging technique that can track microscopic displacements over a field of view that spans several centimetres.

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**, 11 (2021)

doi: <https://doi.org/10.1038/d41586-021-02301-3>

References

1. 1.

[Phys. Rev. Lett. \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02301-3>

- RESEARCH HIGHLIGHT
- 25 August 2021

Why gassy planets are bigger around more-massive stars

During formation, the planets can draw a greater amount of hydrogen and helium around themselves, causing them to increase in size.



Exoplanet CoRoT-7b (artist's impression) orbits a relatively low-mass star.
Credit: ESO/L. Calçada

Planets that form around more-massive stars can efficiently wrap themselves in a blanket of gas — making them larger than planets around less-massive stars.

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**, 11 (2021)

doi: <https://doi.org/10.1038/d41586-021-02300-4>

References

1. 1.

[Astron. Astrophys. \(2021\)](#)

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02300-4>

News in Focus

- **[Preprint ban, ancient DNA and funder pressure](#)** [01 September 2021]
News Round-Up • The latest science news, in brief.
- **[Afghanistan's terrified scientists predict huge research losses](#)** [27 August 2021]
News • For 20 years, science has blossomed in Afghanistan. Now many researchers are fleeing and those who remain face lost funding and the threat of persecution.
- **[Genetic patterns offer clues to evolution of homosexuality](#)**
[23 August 2021]
News • Massive study finds that genetic markers associated with same-sex encounters might aid reproduction. But some scientists question the conclusions.
- **[Home seismometers provide crucial data on Haiti's quake](#)**
[20 August 2021]
News • A volunteer network helps to monitor aftershocks and illuminate the country's earthquake hazards.
- **[Australia's cane toads evolved as cannibals with frightening speed](#)** [25 August 2021]
News • Study finds that the noxious pests have become so numerous, they've developed a taste for each other — as well as defences to ward off such attacks.
- **[The Venezuelan health-care workers secretly collecting COVID stats](#)** [25 August 2021]
News • Faced with government suppression and limited resources, doctors and nurses are quietly working with research networks to report reliable data.
- **[What's next for lab-grown human embryos?](#)** [31 August 2021]
News Feature • Researchers are now permitted to grow human embryos in the lab for longer than 14 days. Here's what they could learn.

- NEWS ROUND-UP
- 01 September 2021

Preprint ban, ancient DNA and funder pressure

The latest science news, in brief.



The compressed skull and teeth of a young woman were found inside an Indonesian cave. Credit: University of Hasanuddin

Fossil DNA hints at mysterious Toalean people

The 7,000-year-old skeleton of a teenage hunter-gatherer from Sulawesi in Indonesia might be the [first remains found from a mysterious, ancient](#)

culture known as the Toaleans. Sulawesi has some of the world's oldest cave art, but ancient human remains have been scarce there.

The largely complete fossil of a roughly 18-year-old Stone Age woman was found in 2015, buried in a limestone cave.

DNA extracted from the skull suggests that she shared ancestry with New Guineans and Aboriginal Australians, as well as with the extinct Denisovan species of ancient human.

“This is the first time anyone’s found ancient human DNA in that region,” says Adam Brumm, an archaeologist at the Australian Research Centre for Human Evolution at Griffith University in Brisbane, who is part of the team behind the find ([S. Carlhoff *et al. Nature* 596, 543–547; 2021](#)).

The authors say she may be one of the Toalean people, whose existence is known from scant archaeological evidence, such as notched stone tools, and who are thought to have lived in Sulawesi around the same time.

The remains were found with Toalean-type tools, providing strong evidence of the woman’s link to these little-known people, agrees archaeologist Shimona Kealy at the Australian National University in Canberra.



Australian ecologists study the distribution of seagrass at low tide in Coronet Bay, Victoria. Credit: Izzet Noyan Yilmaz/Alamy

Preprint ban deemed ‘plain ludicrous’

Australia’s major research funder has ruled more than 30 fellowship applications ineligible because they mentioned preprints and other non-peer-reviewed materials, sparking an [outcry from scientists](#) who say the move is a blow to open science and will stymie careers.

At a time when the COVID-19 pandemic has brought the use of preprints to the fore, researchers say the stance by the Australian Research Council (ARC) — which limits applicants’ ability to refer to the latest research — is out of step with modern publishing practices and at odds with overseas funding agencies that allow or encourage the use of preprints.

Researchers have taken to Twitter in outrage, calling the blanket ruling “short sighted”, “plain ludicrous”, “cruel”, “astonishing”, “outdated” and “gut-wrenching”.

Nick Enfield, a linguistic anthropologist at the University of Sydney, who is currently funded by the ARC, argues that the decision is unconscionable and unethical. “The leading research-funding body of the country is potentially throwing away valuable research on a ridiculous technicality,” he says.

The ARC did not answer specific questions from *Nature* about its rationale for excluding preprints, or confirm how many applicants had been deemed ineligible as a result, but a spokesperson said that the rule “ensures that all applications are treated the same”, adding that “eligibility issues may arise in a number of ways”.

In tweets on 30 August, the ARC responded to the complaints, saying it had “commenced a rapid review” of its policy.



Public health studies have probed the link between fast food and problems such as obesity, high cholesterol and depression. Credit: Mike Kemp/In Pictures/Getty

Funders pressure researchers to suppress results

A survey of public-health researchers has found numerous [instances of trial results being suppressed](#) on topics such as nutrition, sexual health, physical activity and substance use, with 18% of respondents reporting that they had, on at least one occasion, felt pressured by funders to delay reporting, or to alter or not publish findings.

The survey involved 104 researchers from regions including North America, Europe and Oceania who have led trials to evaluate behavioural interventions designed to improve public-health outcomes ([S. McCrabb et al. PLoS ONE 16, e0255704; 2021](#)). These trials, published between 2007 and 2017, were cited in Cochrane reviews, considered the ‘gold standard’ of evidence used to inform health-care decision-making.

Public-health research has a history of industry interference, so the authors, led by Sam McCrabb at the University of Newcastle in Australia, expected industry-funded studies to be those most affected. “But we didn’t find any instances of that,” she says.

In the survey, trial investigators were asked if they had encountered suppression, ranging from requests to change methods or alter conclusions through to appeals to delay publication or not release results.

The authors found that respondents were most likely to report pressure from government-department funders seeking to influence research outcomes.

HOW TRIAL FINDINGS WERE SUPPRESSED

A survey of 104 public-health researchers from around the world found a variety of ways in which funders attempted to interfere in the publication of results of health-intervention trials.



©nature

Source: McCrabb, S. *et al.* *PLoS ONE* **16**, e0255704 (2021).

Nature **597**, 13 (2021)

doi: <https://doi.org/10.1038/d41586-021-02342-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02342-8>

- NEWS
- 27 August 2021

Afghanistan's terrified scientists predict huge research losses

For 20 years, science has blossomed in Afghanistan. Now many researchers are fleeing and those who remain face lost funding and the threat of persecution.

- [Smriti Mallapaty](#)



Afghans who hope to be evacuated head through flooded streets towards Kabul's airport.Credit: Marcus Yam/Los Angeles Times/Shutterstock

On Sunday 15 August, geologist Hamidullah Waizy was interviewing job candidates at the Ministry of Mines and Petroleum in Kabul when he was told the Taliban had entered the city, and he must evacuate. The next morning, he saw armed militants on the streets.

Waizy, a researcher at Kabul Polytechnic University who was recently also appointed director-general of prospecting and exploration of mines at the ministry, was shocked by the city's rapid fall. Since then, he's lived in limbo, mostly shuttered up in the relative safety of his home.

Across Kabul, most universities and public offices remain closed. The Taliban says it wants officials to continue working, but it is not clear what this will look like. "The future is very uncertain," Waizy told *Nature*.

How US sanctions are crippling science in Iran

When the fundamentalist group last held the country, in 1996–2001, it brutally enforced a conservative version of Islamic Sharia law, characterized by women's-rights violations and suppression of freedom of expression. But after it was overthrown in 2001, international funding poured into Afghanistan and universities thrived.

Now, academics fear for their own safety. They also worry that research will languish without money and personal freedoms, and because educated people will flee. Some fear that they could be persecuted for being involved in international collaborations, or because of their fields of study or their ethnicity.

Hard-earned gains

"The achievements we had over the past 20 years are all at great risk," says Attaullah Ahmadi, a public-health scientist at Kateb University in Kabul.

According to news reports, billions of dollars in overseas finance for Afghanistan's government — such as assets held by the US Federal Reserve and credit from the International Monetary Fund — have been frozen. It's not clear whether or when the funding will be released, and how that will affect universities and researchers, but many report salaries not being paid.

In 2001, after the September 11 terrorist attacks in the United States, a US-led coalition invaded Afghanistan and overthrew the Taliban. In 2004, a new government was elected.

Kenneth Holland, a dean at O.P. Jindal Global University in Sonipat, India, was president of the American University of Afghanistan (AUAF) in Kabul in 2017–19. He says that when he arrived in the country in 2006, he found “almost no research being done at universities; no culture of research”.

Since 2004, the World Bank, the US Agency for International Development and other international organizations have poured hundreds of millions of dollars into universities to support teaching, faculty training and some research, he says.

Some three dozen public universities have been established or re-established since 2010, and tens more private universities have been set up. Public universities are funded by the Ministry of Higher Education, which is financed by international donors, says Holland. Private universities survive on tuition fees, although the AUAF is mostly funded by the US government.



Students graduate at the American University of Afghanistan in 2019.Credit: Kiana Hayeri/The New York Times/Redux/eyevine

Hopes and aspirations

The student population at public universities grew from 8,000 in 2001¹ to 170,000 in 2018, one-quarter of whom by that time were women². And although Afghanistan's contribution to international journals remained small, the number of papers recorded annually in the Scopus database increased from 71 in 2011 to 285 in 2019.²

Shakardokht Jafari, a medical physicist at the University of Surrey in Guildford, UK, who is originally from Afghanistan, has seen much progress since 2001, from burgeoning enrolment of female students to growing output on topics from cancer to geology. But now she fears “there will be a stagnation of science and research progress”.

For a long time “scientists considered Afghanistan a black hole”, says Najibullah Kakar, a geohazards scientist at the GFZ German Research Centre for Geosciences in Potsdam. He is one of many Afghans who went abroad for their education, intending to return with new skills to help build the nation. In 2014, he helped to install Afghanistan’s first seismic network to study plate tectonics. He continued that work until 2019, when conflicts made it difficult to travel to remote areas.

He and his team planned to establish a seismic monitoring and research centre in Afghanistan to warn of natural hazards. But since the fall of Kabul, they have been in a state of panic, and Kakar, who says he has not slept for days, is desperately trying to help get his colleagues out.



Taliban fighters patrol the streets in Kabul. Credit: Marcus Yam/Los Angeles Times/Getty

Scholars under threat

Kakar's colleagues are among a tide of researchers seeking asylum overseas. Rose Anderson, a director at humanitarian organization Scholars at Risk (SAR) in New York City, which finds threatened scholars safe havens at universities, says that in August alone, SAR received more than 500 applications from people in Afghanistan.

Some are law scholars who fear reprisals if their field is at odds with the Taliban's interpretation of Sharia law. Many women fear being targeted for their gender and women's-rights activism; some men fear being punished for teaching or supervising women. Others worry that they could be added to hit lists because they studied abroad or have international connections.

Almost all “reported a fear of being targeted just because they are in favour of free and critical inquiry and held ideals around respect for human rights

and women's rights", says Anderson. Many have gone into hiding, or plan to cross into neighbouring countries.

So far, Anderson says, 164 institutions globally have agreed to host scholars, and SAR has appealed to US and European governments to fast-track visas and continue evacuation flights.

But getting people out is difficult: embassies are closed, Kabul airport is overrun and dangerous to get to, and escape over land is hard. Many at risk remain in Afghanistan.

Holland says researchers at the AUAF are particularly vulnerable. The institution has been attacked by militants before: in 2016, 13 people were killed, including faculty members, staff and students. All 60 or so non-Afghan staff members have been evacuated, but only about 20 of some 400 local employees have been flown out, he says. Another 800 students and more than 1,000 alumni could become targets, Holland says.

Risk to minority groups

The largest share of Afghanistan's population of 39 million, including many members of the Taliban, is ethnically Pashtun. Researchers from other ethnic groups risk persecution.

Musa Joya is a medical physicist at Tehran University of Medical Science in Iran, who also works as a lecturer in Kabul. He belongs to the Farsi-speaking Hazara community, which he says makes him a target. He had planned to return to Kabul next year to work at a radiotherapy centre supported by the International Atomic Energy Agency, but those plans could be suspended. Remaining in Iran might not be a solution either, because it is difficult for non-nationals to gain employment at research institutes, Joya says.



[India–Pakistan nuclear escalation: where could it lead?](#)

His wife and children are still in Afghanistan. “I really see a dark future,” he says. “I don’t know how to feed my family; how to rescue them; how to protect them.”

He hasn’t heard reports of the Taliban persecuting people in Kabul, but news of killings in other provinces alarms him. People are “getting ready for a storm”, he says.

There are a few hints that things might not be as restrictive as they were under the previous Taliban administration. Several researchers report that the Taliban is in discussion with university heads about restarting classes. There are suggestions that women might be allowed to continue their studies, although the Taliban has ordered that women and men be taught separately, and some universities have proposed introducing partitions in classrooms.

But in the city of Bamyan, west of Kabul, women have been told not to work and to stay at home, says a female lecturer and education researcher there, who graduated from AUAF and is Hazara. “I am under threat from the Taliban now,” she says.

Appeals for support

Scientists also worry about the future of research. Joya fears that the Taliban won’t prioritize research, or recognize its value. And he does not know how universities will cope without international financial support.

One Kabul-based scholar and member of the Afghanistan Science Academy, who does not want to be named, says this is the third time that he and his family, like many in Afghanistan, have lost everything. He fled during the unrest in the late 1970s ahead of the invasion of the Soviet Union; again in the late 1990s during the Taliban's previous tenure; and is now considering fleeing once more. "It is a very difficult situation for a human being: you are born in war, you grow up in war and now you will die in war."



[US civil-rights group offers support to researchers facing China scrutiny](#)

Many people with postgraduate degrees have already fled. "This is a big catastrophe for the future of Afghanistan," he says. "There will be no educated people left."

The academy, for example, employed some 200 scholars and 160 other staff, with an annual budget of some 300 million afghani (US\$3.5 million), he adds. But they, and many government employees, have not been paid for two months, as the Taliban tightened its grip on the nation.

"The system is almost paralysed," says Ahmadi.

It is not clear whether the international community will recognize the new government and continue to provide funding. Researchers hope they will not be abandoned. "We spent all our money, energy and time in Afghanistan to

build a brighter future for ourselves and our children. But with this kind of withdrawal, they destroyed all our lives, all our hopes and ambitions," says Joya.

Nature **597**, 15-16 (2021)

doi: <https://doi.org/10.1038/d41586-021-02353-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02353-5>

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

- NEWS
- 23 August 2021

Genetic patterns offer clues to evolution of homosexuality

Massive study finds that genetic markers associated with same-sex encounters might aid reproduction. But some scientists question the conclusions.

- [Sara Reardon](#)



Same-sex attraction seems to be at least partly controlled by genetics. Credit: Piotr Lapinski/NurPhoto via Getty

To evolutionary biologists, the genetics of homosexuality seems like a paradox. In theory, humans and other animals who are exclusively attracted to others of the same sex should be unlikely to produce many biological children, so any genes that predispose people to homosexuality would rarely be passed on to future generations. Yet same-sex attraction is widespread in humans, and research suggests that it is partly genetic.

In a study of data from hundreds of thousands of people, researchers have now identified genetic patterns that could be associated with homosexual behaviour, and showed how these might also help people to find different-sex mates, and reproduce. The authors say their findings, published on 23 August in *Nature Human Behaviour*¹, could help to explain why genes that predispose people to homosexuality continue to be passed down. But other scientists question whether these data can provide definitive conclusions.



No ‘gay gene’: Massive study homes in on genetic basis of human sexuality

Evolutionary geneticist Brendan Zietsch at the University of Queensland in Brisbane, Australia, and his colleagues used data from the UK Biobank, the US National Longitudinal Study of Adolescent to Adult Health and the company 23andMe, based in Sunnyvale, California, which sequence genomes and use questionnaires to collect information from their participants. The team analysed the genomes of 477,522 people who said they had had sex at least once with someone of the same sex, then compared

these genomes with those of 358,426 people who said they'd only had heterosexual sex. The study looked only at biological sex, not gender, and excluded participants whose gender and sex did not match.

In earlier research, the researchers had found that people who'd had at least one same-sex partner tended to share patterns of small genetic differences scattered throughout the genome². None of these variations seemed to greatly affect sexual behaviour on its own, backing up previous research that has found [no sign of a ‘gay gene’](#). But the collection of variants seemed to have a small effect overall, explaining between 8% and 25% of heritability.

Next, the researchers used a computer algorithm to simulate human evolution over 60 generations. They found that the array of genetic variations associated with same-sex behaviour would have eventually disappeared, unless it somehow helped people to survive or reproduce.

Overlapping genes

Zietsch and his team decided to test whether these genetic patterns might provide an evolutionary edge by increasing a person's number of sexual partners. They sorted the participants who had only had heterosexual sex by the number of partners they said they had had, and found that those with numerous partners tended to share some of the markers that the team had found in people who had had a same-sex partner.

The researchers also found that people who'd had same-sex encounters shared genetic markers with people who described themselves as risk-taking and open to new experiences. And there was a small overlap between heterosexual people who had genes linked to same-sex behaviour and those whom interviewers rated as physically attractive. Zietsch suggests that traits such as charisma and sex drive could also share genes that overlap with same-sex behaviour, but he says that those traits were not included in the data, so “we’re just guessing”.



[Sexual-minority students more likely to abandon science majors](#)

The authors acknowledge many limitations of the study. All of the participants lived in the United Kingdom or United States, and were of European descent. And the databases' questionnaires asked about sexual behaviour, not sexual attraction. Most of the participants were born during a time when homosexuality was either illegal or culturally taboo in their countries, so many people who were attracted to others of the same sex might never have actually acted on their attraction, and could therefore have ended up in the wrong group in the study.

Julia Monk, an ecologist and evolutionary biologist at Yale University in New Haven, Connecticut, thinks that these caveats are so important that the paper can't draw any real conclusions about genetics and sexual orientation. Sexual behaviour and reproduction, she says, occupy a different place in modern societies than they did for human ancestors, so it's difficult to infer their role in our evolution. For instance, people might engage with more sexual partners now that sexually transmitted diseases can be cured. And the existence of birth control and fertility treatments negates many of the reproductive advantages that genes might provide. "It's clear that people's behaviour when it comes to sex and reproduction is highly culturally informed, and maybe digging into genetics is next to impossible," Monk says.

Weak linkages

Qazi Rahman, a psychologist at King's College London, thinks that the study was well-conducted, but he is sceptical of some of the conclusions. He says the data sets are too biased towards people who were willing to reveal their sexual behaviour to researchers, which could itself be considered a risky behaviour that could be reflected in the genetic data. He adds that once the data are broken down into men and women, and into those who had only had same-sex partners versus those who had encounters across sexes, the number of people in each group becomes so small that the genetic linkages are very weak.



Don't ignore genetic data from minority populations

Dean Hamer, a retired geneticist in Haleiwa, Hawaii, who published some of the first studies on the genetics of sexual orientation, is disappointed with the study. Defining sexual orientation on the basis of a single same-sex encounter is not a useful way of categorizing people, he says, because many people who identify as heterosexual have experimented with a same-sex partner. "You're not even asking the right people the right question," Hamer says. Instead, he thinks the researchers have found genetic markers associated with openness to new experiences, which could explain the

overlap between people who have had a homosexual partner and heterosexual people who have had many partners.

Zietsch says that risk-taking can explain only part of the statistical link between markers associated with same-sex encounters and those associated with number of partners. And he admits that using a single homosexual experience as an indication of sexual orientation isn't ideal, but says that the UK Biobank didn't provide data on attraction. Zietsch's previous research on data from 23andMe showed a strong genetic overlap between people who reported same-sex sexual experiences and those who reported same-sex attraction, suggesting that the same genes controlled both factors.

Hamer acknowledges that linking a complex behaviour to genetics is extremely difficult, but says he is glad the team is researching sexual orientation. "It's vastly understudied considering it's a driving force for the human race," he says. "It's a good question, they just didn't find an answer."

Nature **597**, 17-18 (2021)

doi: <https://doi.org/10.1038/d41586-021-02312-0>

References

1. 1.

Zietsch, B. P. *et al. Nature Hum. Behav.*
<https://doi.org/10.1038/s41562-021-01168-8> (2021).

2. 2.

Ganna, A. *et al. Science* **365**, eaat7693 (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02312-0>

- NEWS
- 20 August 2021

Home seismometers provide crucial data on Haiti's quake

A volunteer network helps to monitor aftershocks and illuminate the country's earthquake hazards.

- [Alexandra Witze](#)



The earthquake in Haiti this month has destroyed many buildings, such as the Church St. Anne in Chardonnières, shown here. Credit: Reginald Louissaint Jr/AFP via Getty

A network of inexpensive seismometers, installed in people's living rooms, gardens and workplaces across Haiti, is helping scientists to unravel the inner workings of the magnitude-7.2 earthquake that devastated the southwestern part of the Caribbean nation this month. The community-science effort launched after the country's last major earthquake — a magnitude-7 tremor in 2010 that killed more than 100,000 people — and has since helped to reveal details about Haiti's seismic activity.

In a country whose official seismic-monitoring stations are sometimes offline because of limited resources, the community-seismology project provides much-needed data. Right now, the network is detecting aftershocks that continue to rattle the region. Its seismometers feed data into a system that displays the locations and magnitudes of Haitian earthquakes on a [web-based portal](#) in real time.

"It's not professional equipment, and there are a lot of limitations," says Dominique Boisson, a geologist at the State University of Haiti in Port-au-Prince who helps to run the network. But "some results are very nice".

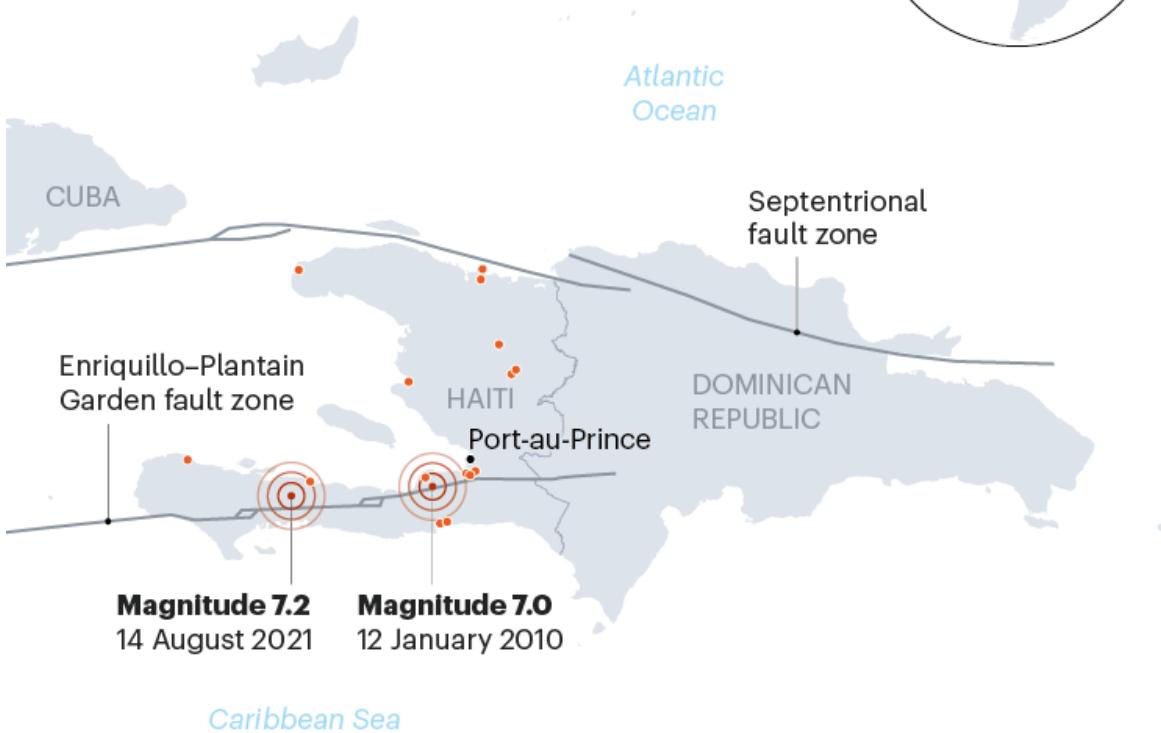
Difficult work

The network underscores just how far seismology in Haiti has come in 11 years. When the 2010 earthquake struck near Port-au-Prince, the country had no seismologists and just one official seismic-monitoring station, says Boisson. Now, [there are several professional seismologists](#), as well as 7 stations in the official national network, which is operated by Haiti's Bureau of Mines and Energy, and 15 in the community-science network.

TRACKING HAITIAN TREMORS

Low-cost seismometers hosted in people's homes and workplaces are contributing to the scientific understanding of the recent earthquake in Haiti. The community-based science effort launched after a massive quake in 2010 that killed many.

- Volunteer-science stations



©nature

Sources: C. S. Prentice *et al.* [*Nature Geosci.*](#) 3, 789–793 (2010)/Ayiti-Séismes

Within days of the big quake hitting on 14 August, teams of scientists and technicians were driving towards its epicentre, carrying seismometers and other instruments to measure how the ground was moving. Monitoring the Earth with scientific instruments immediately after a quake allows researchers to better understand why the earthquake occurred and the future seismic risk. In 2010, it took weeks after the quake for foreign researchers to fly to Haiti and deploy instruments.

This year, many of those foreign teams are forbidden to travel to Haiti because of COVID-19 restrictions and political instability following the assassination in July of Haiti's president, Jovenel Moïse. Instead, the work is

being led by Haitian seismologists, such as Steeve Symithe, also at the State University, who, before he went into the field, [was streaming Facebook Live presentations](#) about the science of the quake to the Haitian public.

Both the 2010 and 2021 quakes happened in the Enriquillo–Plantain Garden fault zone, a tangle of fractures in Earth’s crust where the North American and Caribbean tectonic plates slide past one another. It runs from west to east along Haiti’s southern peninsula. The 2010 quake occurred on [a previously unknown fault in that zone](#). The epicentre of the 2021 quake lies about 100 kilometres to the west, in the province of Nippes.

At least 2,100 people died in the 14 August quake, although the total count has yet to be tallied. The US Geological Survey estimates that there might have been more than 10,000 deaths. Many survivors endured heavy winds and rain from a tropical storm as they tried to shelter outside. The scientists en route to the area spent the night in their cars as rain pelted down, softening the ground and generating landslides as aftershocks shook the ground, Boisson tells *Nature*. “It was pretty difficult” for them, he says.

DIY seismology

The challenge of doing fieldwork in Haiti helped to inspire the creation of the community-seismology project in 2019. That was when Eric Calais, a seismologist at the École Normale Supérieure in Paris who has studied Haiti’s earthquakes for years, happened across a company that sells seismic stations to hobbyists. Looking for ways around the national Haitian network’s intermittent data, he used leftover money from a grant to buy some stations. Known as Raspberry Shakes, they contain tiny accelerometers that detect when the ground trembles and send that information to be processed and commingled with that from other stations.



Earthquake-risk maps pinpoint world's most vulnerable areas

These US\$500 stations are not as sophisticated as Haiti's official \$50,000 monitoring stations. "But when it comes to locating quakes, determining magnitude, doing basic seismology — they are really excellent," says Calais. And because they are in people's homes and workplaces, they more often have a steady supply of power and reliable Internet access. The team, which includes Calais, Boisson, Symithe and many others, recruited people to host the stations. Boisson had one in his garden until last week, when he dismantled it to move it closer to the epicentre of the 14 August quake. The host who had the Raspberry Shake closest to the epicentre was chagrined that his station was offline during the quake; he immediately ran out and topped up his Internet plan, says Calais, and the station was soon back up and running.

Funded by international supporters, Calais and his colleagues have kept the network of 15 stations operational for two years¹; they aim to soon ramp up to 50 or more stations. Community-seismology networks have sprung up in other places around the world, but the Haiti network is unique in providing data in an area where few seismic data are otherwise collected, says Calais.

The Haitian community-seismology data feed into a nationwide experimental system called Ayiti-Séismes, which is hosted at a website run by the Côte d'Azur University in Nice, France. Ayiti-Séismes also pulls data

from official seismic stations in Haiti as well as those in nearby countries, including the Dominican Republic and Cuba. The result is a real-time map of aftershocks blanketing southwestern Haiti in shades of red and orange. “The network is alive and well,” says Susan Hough, a seismologist at the US Geological Survey in Pasadena, California, who has worked in Haiti for many years, including after the 2010 quake.

Future risk

The quake’s epicentre is fairly close to quakes that occurred in 1952 and 1953, which were probably between magnitude 5 and 6, says Calais. In terms of future risk, the Enriquillo–Plantain Garden fault zone could still yield another major quake. “In this area, we cannot say that it’s over,” says Boisson. Some speculate that the 2010 quake contributed to the recent one by transferring stress towards the region that just ruptured — and that seismic risk remains high in Port-au-Prince and across much of the Enriquillo–Plantain Garden fault zone.



[Artificial intelligence nails predictions of earthquake aftershocks](#)

Boisson notes that many scientists have been worried about a different major geological region in Haiti’s north, known as the Septentrional fault zone; it unleashed a major quake in 1842. “After 2010, we thought it would be this

fault” that would cause future quakes, he says. “And then it was in the south again.”

About 600 aftershocks have been detected from the 14 August quake so far — compared with roughly 10 in the same time period after the 2010 quake, although there were undoubtedly more that were not detected, says Calais. “We now have very strong information about not only where the [14 August] quake occurred, but also how wide the rupture was, in which direction the fault was dipping,” he adds. “That’s essential” to understanding why the quake occurred and what to expect in the future.

Nature **597**, 18-19 (2021)

doi: <https://doi.org/10.1038/d41586-021-02279-y>

References

1. 1.

Calais, E. *et al. Front. Earth Sci.*
<https://doi.org/10.3389/feart.2020.542654> (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02279-y>

- NEWS
- 25 August 2021

Australia's cane toads evolved as cannibals with frightening speed

Study finds that the noxious pests have become so numerous, they've developed a taste for each other — as well as defences to ward off such attacks.

- [Max Kozlov](#)



The cane toad, whose skin is toxic, has devastated the populations of some animals in Australia, where it is an invasive species. Credit: Ian Waldie/Getty

The list of ‘deadly animals in Australia’ just got a little weirder. The cane toad, a toxic, invasive species notorious for devouring anything it can fit in its mouth — household rubbish, small rodents and even birds — has become highly cannibalistic in the 86 years since it was introduced to the continent, according to a new study. Its counterpart in South America, where cane toads originated, is far less cannibalistic.

The discovery could help researchers to understand the evolutionary underpinnings of how this uncommon and extreme behaviour emerges. Scientists have seen cannibalism evolve in species before, says Volker Rudolf, a community ecologist at Rice University in Texas, who studies the phenomenon. But what’s exciting about this work, he says, is that the researchers are almost seeing it “develop in front of their eyes”, given that the behaviour arose in less than a hundred years — the blink of an eye by evolutionary standards.

“These toads have gotten to the point where their own worst enemy is themselves,” says Jayna DeVore, an invasive-species biologist at Tetiaroa Society, a non-profit organization in French Polynesia, and a co-author of the study, which was published on 23 August in *Proceedings of the National Academy of Sciences of the United States of America*¹. Scientists estimate that there are well over 200 million of the amphibians in Australia. They have become so abundant, says DeVore, that they face more evolutionary pressure from each other, as they compete for resources, than from anything else in Australia.

Tadpole terror

Farmers first introduced about 100 cane toads (*Rhinella marina*) to Australia from their native range in South America in 1935 to control cane beetles (*Dermolepida albohirtum*), which were wreaking havoc on sugarcane plantations. The giant toads failed to knock down the beetle populations, but they succeeded in epically multiplying. Because of their highly poisonous skin, which is coated in bufotoxins, they had no natural predators and went on to invade large swaths of the northern and eastern parts of the country.

Although adult cane toads are fearsome — they grow up to 25 centimetres in length — it's their tadpoles that are usually the cannibals. Multiple tadpoles together can gobble more than 99% of the hatchlings that emerge from the tens of thousands of eggs in a single clutch².



Cane-toad tadpoles (bottom) in Australia can work together to devour most of the hatchlings of their own kind that emerge from a single clutch of eggs (top). Credit: Anton Sorokin/Alamy

DeVore and colleagues were curious to see whether the cannibalistic behaviour was common across all cane toads, or if it was due to how invasive the Australian ones are. So they collected cane toads from Australia and from French Guiana, and bred them to produce hatchlings and older tadpoles. The team then exposed a single tadpole to 10 hatchlings from its group — either from Australia or South America — hundreds of times and found that invasive Australian tadpoles were 2.6 times as likely to cannibalize hatchlings as native South American ones.

Researchers have long known that the Australian tadpoles are attracted to the hatchlings because of the scent of the younger animals' toxic skin.

“You’ll get this huge avalanche of thousands and thousands of tiny cane-toad tadpoles coming toward this chemical,” says Rick Shine, an evolutionary biologist at the University of Sydney in Australia, and a co-author of the study. DeVore, Shine and co-workers saw this play out in their experiments: the Australian tadpoles were nearly 30 times as likely to swim towards a trap containing hatchlings as an empty trap, and the South American tadpoles showed no preference for either.



[Ecologists try to speed up evolution to save Australian marsupial from toxic toads](#)

Although the speed with which the toads evolved this behaviour is impressive, the team was even more surprised by how fast the animals evolved a defence to protect against it. The researchers found that when invasive Australian hatchlings shared a tank with caged, older tadpoles from the same group, the hatchlings were more likely to have a shorter developmental period than that of the South American hatchlings. Older tadpoles don’t tend to eat older tadpoles — so the toads might have evolved to speed up their hatchling phase, the researchers found. This would limit the amount of time they spend vulnerable to cannibalism, even if the adaptation eventually stunts their growth, says DeVore.

Roshan Vijendravarma, an evolutionary biologist at the Curie Institute in Paris, who has studied cannibalism in fruit flies, says the differences

between the invasive and native toads' behaviour probably have a genetic basis, given how extreme they are and how quickly they evolved over relatively few generations of toads.

Shine and his colleagues think this idea is worth exploring and are studying it now. Although there are still mysteries around the cane toads' cannibalistic tendencies, one thing is for certain, says Shine: "The cane toads that are currently hopping across Australia are extraordinarily different animals from the ones that were first taken out of the native range."

Nature **597**, 19-20 (2021)

doi: <https://doi.org/10.1038/d41586-021-02317-9>

References

1. 1.

DeVore, J. L. *et al.* *Proc. Natl. Acad. Sci. USA* **118**, e2100765118 (2021).

2. 2.

DeVore, J. L., Crossland, M. R., and R. Shine. *Ecol. Monogr.* **91**, e01426 (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02317-9>

- NEWS
- 25 August 2021

The Venezuelan health-care workers secretly collecting COVID stats

Faced with government suppression and limited resources, doctors and nurses are quietly working with research networks to report reliable data.

- [Luke Taylor](#)



Health-care workers in Venezuela protest against a lack of supplies and for fair salaries amid the COVID-19 pandemic. Credit: Yuri Cortez/AFP via

Getty

“The hardest part of watching my colleagues and patients suffer is not being able to say anything about it,” says Gabriel Romero, an attending physician at one of Venezuela’s largest public clinics.

Romero, who asked that *Nature* not use his real name for fear of punishment by the Venezuelan government, is one of the many health-care workers in the country fighting COVID-19 despite a lack of basic medical equipment, a steady power and water supply, and adequate pay. When they have [spoken out about what they say are unacceptable conditions](#), some have been detained by government forces.

Included in their list of complaints is a lack of reliable COVID-19 data.



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

Officially, Venezuela reports that it has had more than 326,000 COVID-19 infections and about 3,900 deaths amongst its 29 million inhabitants since the start of the pandemic. That’s 135 fatalities per million people. By contrast, its next-door neighbours Colombia and Brazil report about 2,440 and 2,700 fatalities per million people, respectively (see ‘Questionable COVID data’).

In looking at the data for South America, Romero and others say it’s obvious that Venezuela’s numbers do not reflect reality, and that the drastic

undercount is driven by a lack of testing and infrastructure — but also by a deliberate effort by the government to downplay the pandemic. Because accurate statistics are crucial for aid organizations deciding where to send resources, and for local officials mulling whether to open schools and businesses, many doctors and nurses have taken matters into their own hands, collecting data from hospitals and reporting the numbers secretly to various research networks and non-governmental organizations (NGOs). These numbers are much higher than those appearing in official government reports.

“It’s a lot of pressure,” says Romero, who coordinates a clandestine network that collects these data. “I always worry that we could get detained. But I cannot live in a country where the official narrative is that everything is fine when we are living a totally different reality.”

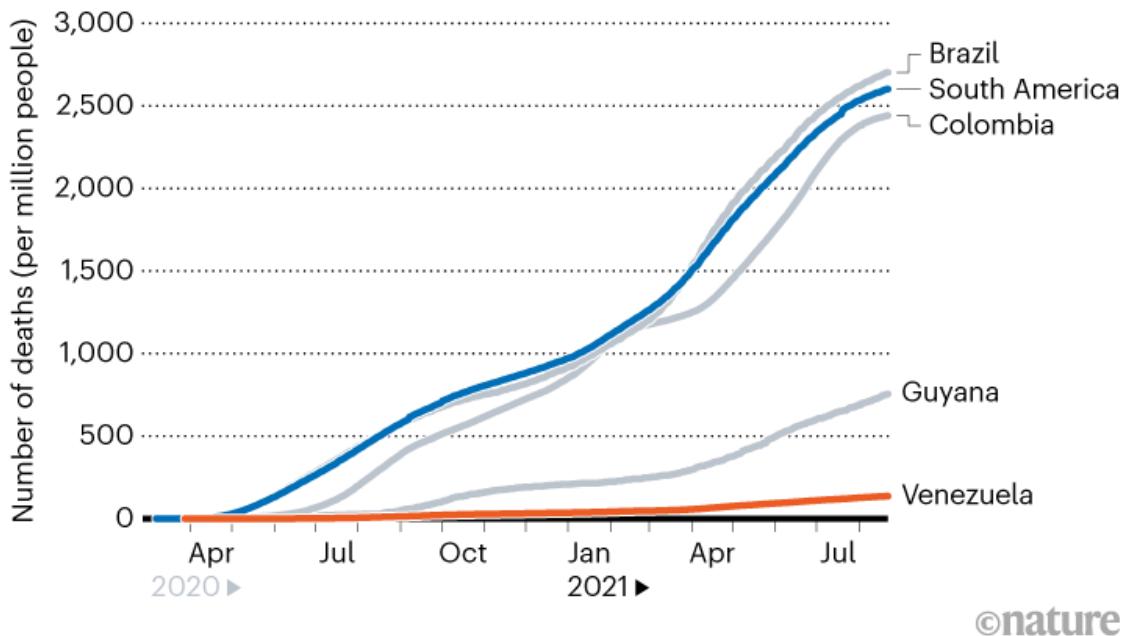
The pandemic narrative

Venezuela’s economy has collapsed in the past decade because of corruption, financial mismanagement and the price of oil, its key export, plummeting. At least 5.4 million people have fled the country as a result, and of those remaining, about 90% now live in poverty, according to the United Nations.

It’s possible that this economic collapse has slowed the spread of the coronavirus SARS-CoV-2, say epidemiologists interviewed by *Nature*. It has made travel across the country difficult, and few Venezuelans have the means to go out to restaurants or bars, where super-spreader events might occur.

QUESTIONABLE COVID DATA

Venezuela's government has reported far fewer cumulative deaths due to COVID-19 than the three nations that border it, when adjusted for population – and is an outlier in South America. Epidemiologists say the drastically lower number signals inaccurate counting.



©nature

Source: Our World in Data

But the government data still do not line up with reports collected at hospitals, they point out. According to Médicos Unidos Venezuela, a collective of Venezuelan doctors monitoring the situation, more than 736 Venezuelan health-care workers have died from COVID-19 since last year. The collective sometimes reports higher weekly death rates for Venezuela's health-care workers than the government reports for the entire country.

"Looking at all these indicators, the government figures are impossible," says María Eugenia Grillet, an epidemiologist at the Central University of Venezuela in Caracas.

"It's evident that the [government] data are completely misguided, incorrect and unhelpful," says Pedro Delgado, vice-president of the Institute for Healthcare Improvement, an NGO based in Boston, Massachusetts. "The numbers make no sense."

Venezuela's Ministry of Health declined to comment when asked by *Nature* about discrepancies in its data compared with those reported by groups such as Médicos Unidos Venezuela and NGOs.

The picture on the ground

Venezuela's true COVID-19 infection and death statistics are at least five to seven times as high as those the government is reporting, says Julio Castro, an infectious-diseases researcher at the Central University of Venezuela. He makes this estimate on the basis of data submitted to the National Hospital Survey in Venezuela, which he coordinates.

In their spare time and in secret, nurses and doctors such as Romero collect data about patient admissions and deaths at Venezuela's 40 major hospitals to provide to the survey.



A woman with COVID-19 receives oxygen from health-care workers at a field hospital in Venezuela. Credit: Ariana Cubillos/AP/Shutterstock

Castro says one of the reasons the government's official statistics are low is a lack of COVID-19 diagnostic tests in the country. Gold-standard tests for determining whether a person is positive for COVID-19 — polymerase-chain reaction (PCR) tests — can be processed only at two approved laboratories in Caracas, he says. The labs are working at capacity, causing delays for weeks. And what's more, for people outside Caracas, testing centres are unreachable because of fuel shortages, he adds.

Because of the lack of diagnostic tests, doctors and nurses submitting data to the National Hospital Survey instead evaluate whether a person has COVID-19 by assessing whether they have an acute respiratory infection — the main symptom of COVID-19. Health-care workers agree that although it isn't a perfect indicator of COVID-19, it provides a more-trusted picture of reality than the government's statistics, which are based on PCR tests.

Beyond borders

Venezuela isn't the only country to struggle with its COVID-19 statistics. In May, Peru [updated its official death statistics](#) after researchers warned they were not representative of the true situation in the country. After taking into consideration [excess deaths](#) — the number of people who died over a certain period, compared with the number of deaths researchers would expect based on a pre-pandemic baseline — the Peruvian government nearly tripled its official death tally.

Even wealthy nations such as the United States have an excess death count that is 20% higher than the government is reporting, probably owing to a variety of factors, such as misdiagnosis. But researchers especially use this method to get a more accurate view of COVID-19 statistics in places such as Venezuela, where tests are sparse, where many people do not die in hospital so their deaths aren't registered, or where government data are deemed untrustworthy.



Why Uruguay lost control of COVID

But in Venezuela, the method can't be used because the Ministry of Health hasn't published mortality statistics since its last national epidemiology report in 2016.

"It's not just epidemiological data, we don't have data for any indicator", says Grillet, whether it's the economy or public health.

In the absence of baseline data, researchers have been trying to get a handle on the situation in Venezuela from outside its borders. The Vector-borne disease control in Venezuela Network, an international group of researchers that studies infectious diseases in Venezuela, wanted to take nasal samples from people in the country suspected of having COVID-19, search for SARS-CoV-2 and sequence its genome, but was barred last year by the Ministry of Health from doing so. The researchers pivoted, and instead analysed samples in 2020 from some of the Venezuelans who had recently fled the economic crisis and migrated to Colombia, to get a window into Venezuela's pandemic.

The network confirmed for the first time that six coronavirus variants — including two, Beta and Gamma, which are classified as variants of concern — had been circulating in Venezuela¹. By contrast, the government has not released any genomic surveillance data.



How many COVID deaths are acceptable in a post-pandemic world?

Operating outside of Venezuela also means that the team that reported these results can avoid being detained by the Venezuelan government. At least 12 health-care professionals have been detained for speaking out about the situation in Venezuela during the pandemic, [according to the human-rights-focused NGO Amnesty International](#). And last year, Diosdado Cabello, a senior government official in President Nicolás Maduro’s administration, [suggested on national television](#) that the Venezuelan Academy of Physical, Mathematical and Natural Sciences (ACFIMAN) — to which Grillet belongs — should be raided and its scientists arrested for releasing models that predicted an increase in coronavirus infections and deaths.

The government “is trying to keep the numbers low so they don’t give the impression that there is a large, uncontrolled outbreak”, says Castro.

Health-care workers in particular are under a lot of pressure just from doing their jobs, says Margarita Lampo, an ecologist who studies infectious diseases at ACFIMAN and the Venezuelan Institute for Scientific Research in Caracas. Add that stress to the fear of getting fired, or worse, for sharing their data, she adds, and it’s much more difficult. “We are very thankful for what they do. They should know that their work has been so important for our country.”

Nature **597**, 20-21 (2021)

doi: <https://doi.org/10.1038/d41586-021-02276-1>

References

1. 1.

Patiño, L. H. *et al.* *Int. J. Infect. Dis.* **110**, 410–416 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02276-1>

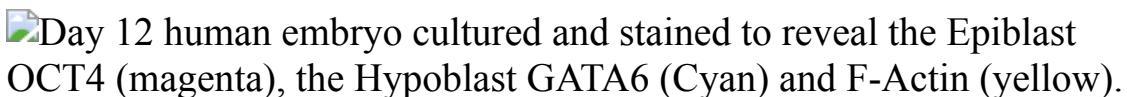
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 31 August 2021

What's next for lab-grown human embryos?

Researchers are now permitted to grow human embryos in the lab for longer than 14 days. Here's what they could learn.

- [Kendall Powell](#) ✉



A human embryo grown in the laboratory for 12 days, showing cells that will form the embryo itself (magenta). Credit: Antonia Weberling, Bailey Weatherbee, Carlos Gantner and Magdalena Zernicka-Goetz

It was day 13 in a set of experiments in Ali Brivanlou's laboratory and he had an agonizing task ahead. His team of developmental biologists had thawed dozens of human embryos, placed them into individual culture dishes and watched them grow through the earliest stages of development — something that only a handful of researchers worldwide had ever seen. But he knew that it had to end.

The embryos would soon bump up against the 14-day rule, an international consensus that human embryos should be cultured and grown in the lab only until 14 days post-fertilization. Day 14 is roughly when the primitive streak appears, a structure that marks the point at which the embryo sets up the body axes, and begins to distinguish head from tail and left from right.

"It was one of the toughest decisions I've had to make in my life, but it was time to stop this experiment," says Brivanlou, recalling the 2015 research in his lab at the Rockefeller University in New York City. On day 13, the team

peeled the embryos from the bottom of the dish and froze them — curtailing any further development. “We didn’t know what was going to happen after that point,” he says.

Now, Brivanlou and other developmental biologists have the chance to find out.

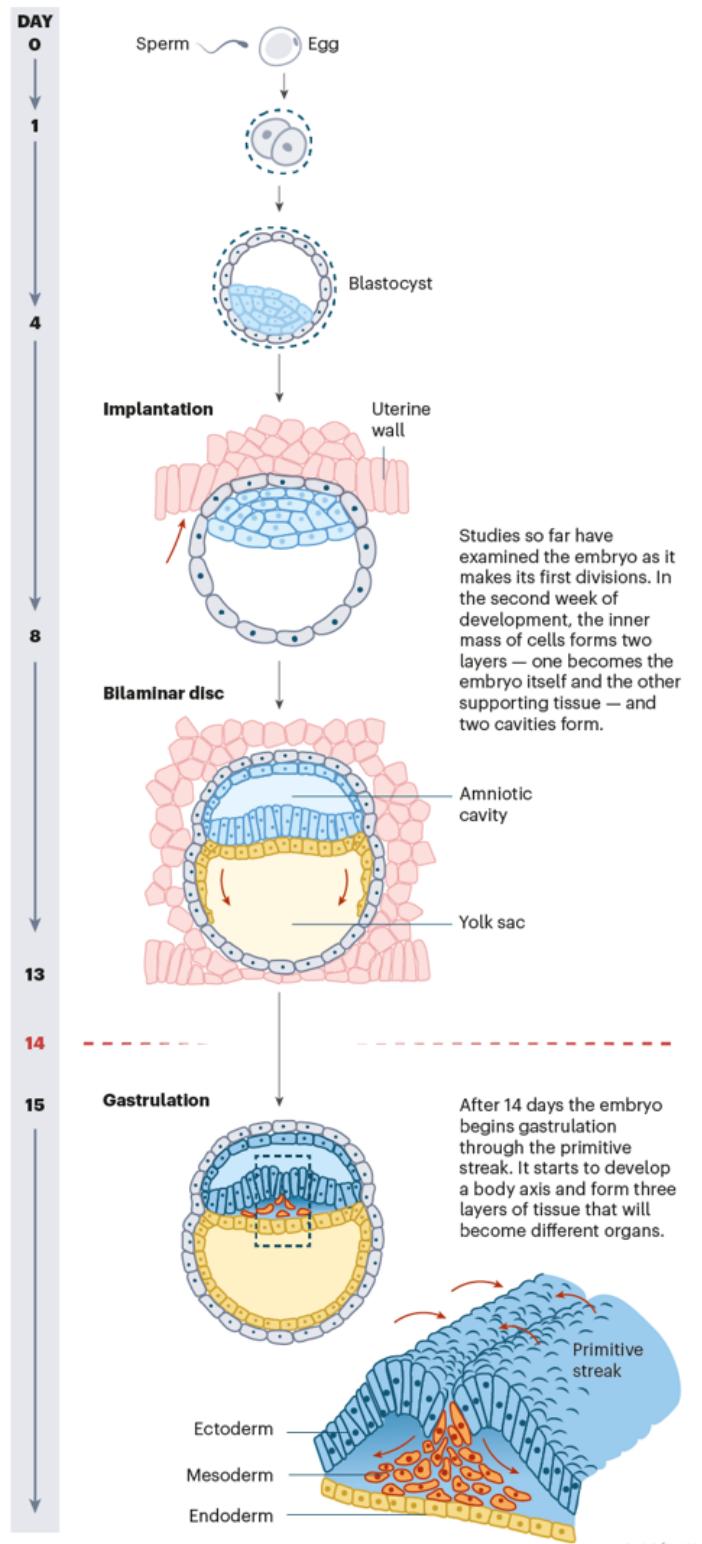
In May, the International Society for Stem Cell Research (ISSCR) released new guidelines¹ that [relaxed the 14-day rule](#), taking away the hard barrier. Although only a few labs around the world have perfected the techniques needed to culture human embryos up to day 14, the science is advancing rapidly. The relaxed rule allows lab groups, in countries where it is legal, to apply to the regulators for permission to continue research past 14 days. Such studies could reveal what happens during human development after the embryo would normally have implanted in the uterus, about a week after fertilization. It then goes through a stage called gastrulation, roughly between days 14 and 22, when the body’s main pattern emerges and the foundations are laid for organ generation.

Cracking open a window on these later stages would allow scientists to better understand the nearly one-third of pregnancy losses and numerous congenital birth defects thought to occur at these points in development. In addition, these stages hold clues to how cells differentiate into tissues and organs, which could boost regenerative medicine.

Most researchers expect that Brivanlou’s group and a smattering of others will immediately push the technical boundary of growing human embryos in the lab for longer stretches. The few groups that can do this are in stiff competition. In other places, laws would have to be changed for such research to move forward (see ‘Lifting the limit’). Some researchers in China, for example, are discussing whether to change their national policy, which currently reflects the 14-day cut-off, says Tianqing Li, an embryologist at Kunming University of Science and Technology.

LIFTING THE LIMIT

So far, researchers have been able to study human embryos until 14 days of development — about a week after they would usually implant into the womb. With the limit lifted, researchers are permitted to explore what happens next.



©nature

Credit: Nik Spencer/*Nature*

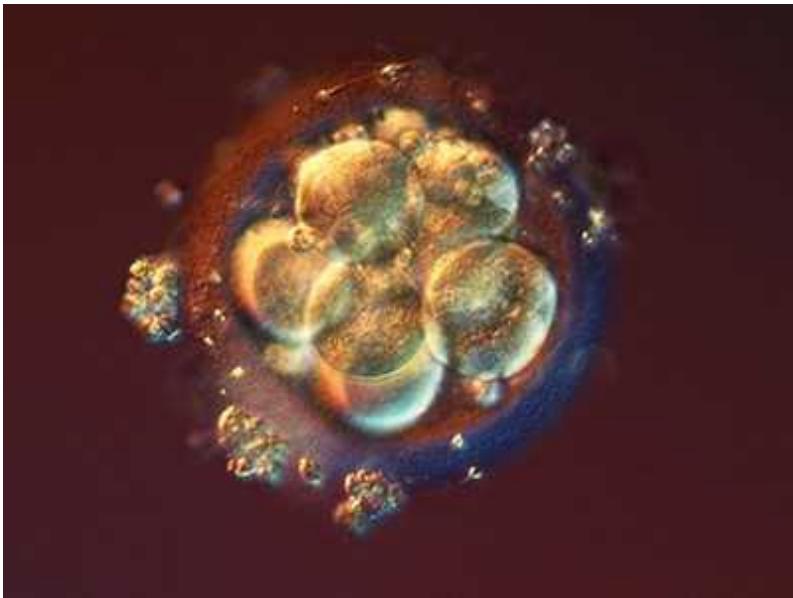
So researchers do not expect a flood of new post-14-day studies just yet — nor are they all convinced such experiments are justified. “I’m cautious about using human embryos as a research system in their own right,” says Naomi Morris, a developmental biologist at the Francis Crick Institute in London.

As a way of skirting the boundary, researchers have in the past five years developed an array of human embryo models, most of which are formed from mixtures of stem cells. These models mimic multiple, but brief, phases of early development and can be made without using the scarce and ethically fraught human embryos donated by people undergoing *in vitro* fertilization (IVF) treatment. So far, the 14-day rule doesn’t apply to these embryo models. But, as they get more sophisticated, with the potential to form recognizable structures, or even organs, they enter their own ethical grey area.

Whether using models or the real thing, scientists say they have a lot to learn. “Embryos are the great masters,” says Nicolas Rivron, a stem-cell biologist and embryologist at the Austrian Academy of Science’s Institute of Molecular Biotechnology in Vienna. “They are the structures that teach us everything about how we are formed and how we fail.”

Pushing the envelope

The 14-day rule was first proposed in 1979, as IVF technology emerged and human embryos existed for the first time outside the body — although the longest they could survive at the time was a few days. By 2006, when the ISSCR issued its first set of guidelines for human embryonic stem cells (ES cells), the 14-day rule was firmly established in the research community.



Limit on lab-grown human embryos dropped by stem-cell body

The guidelines have been widely adopted around the world by researchers and funders. In several countries, including Germany and Austria, it is illegal to do any research on human embryos, and many others impose a 14-day limit by law, such as the United Kingdom, China, Japan, Australia and Canada. In a few places, including the United States and Israel, there are guidelines but no law prohibiting or limiting human embryo research, although it cannot be federally funded in the United States.

The previous update to the ISSCR guidelines came out in 2016, just before two research groups published breakthroughs.

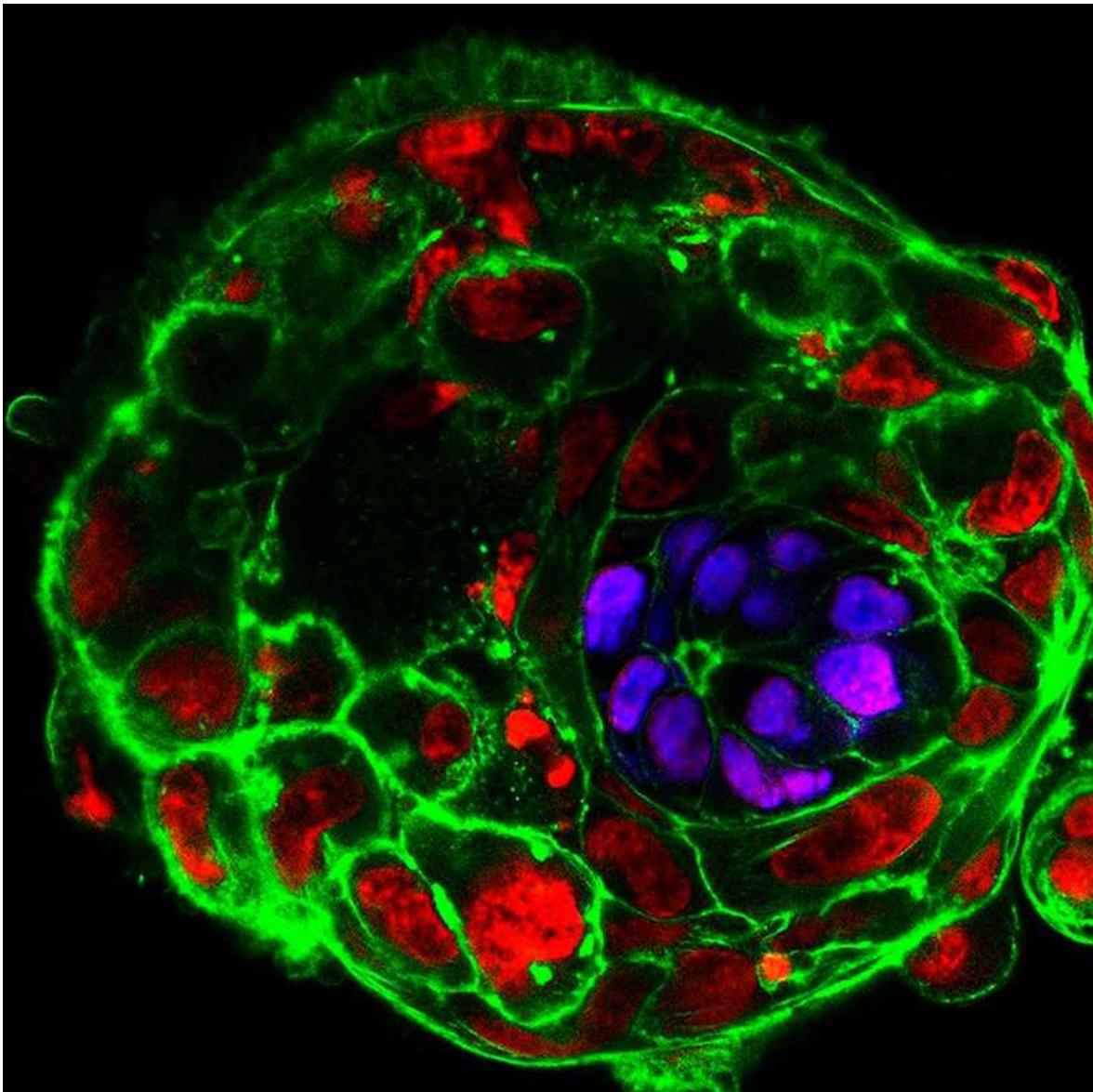
Magdalena Zernicka-Goetz's lab at the University of Cambridge, UK, began the quest to culture human embryos beyond seven days in 2013. The group wanted to understand what happens beyond the blastocyst, or 'ball of cells', stage. Her group finessed the right recipe of hormones and growth factors, and in 2016 her team² and Brivanlou's³ reported that they had grown human embryos until day 12–13.

"The discoveries we are making, such as where humans put our future head, are of major importance," says Zernicka-Goetz, who now splits her time between her lab in the United Kingdom and one at the California Institute of Technology in Pasadena. "I'm fascinated by the second, third and fourth

week of development, which we cannot see with ultrasound, but starts the development of the progenitors for the main organs.”

One question concerns how genes are expressed in the embryo’s cells as it grows. In one of the largest studies of human embryos so far, Zernicka-Goetz’s group analysed 4,820 single cells from 16 embryos developing in the lab from the stage when an embryo would normally implant in the womb (day 5) to the preparations for gastrulation at day 11. Single-cell RNA sequencing revealed which genes switch off and which switch on as embryonic cells transition from totipotency, when they can still become any cell in the body, to pluripotency — a more differentiated state⁴.

Although these developmental signals are known in mice, this study is one of the first to reveal the molecules that underwrite human development.



A lab-grown human embryo ten days after fertilization, showing the cells that will become the fetus in purple. Credit: Zernicka-Goetz Laboratory, Cambridge University/Science Photo Library

In other experiments that can be done only with human embryos in culture, both Brivanlou and Zernicka-Goetz and their teams followed the fate of human embryos with aneuploid cells. These are cells that have an abnormal number of chromosomes, a condition thought to cause up to half of early pregnancy losses.

IVF clinics typically test just a few of an embryo's cells to analyse its genetic health. But experiments suggest that this could be misleading. Zernicka-Goetz's group found that embryos diagnosed with some types of aneuploidy go on to develop normally in the lab⁵. Brivanlou's group have analysed gene expression in human embryos from days 3 to 14, and found that cells with abnormal numbers of chromosomes were eliminated — perhaps by developing into supporting tissues or being culled through cell death⁶. Both studies suggest that the testing for aneuploidy commonly done on IVF embryos probably results in many embryos being wrongly deemed 'unhealthy'.

Working with human embryos is essential for these insights, says Brivanlou. "We cannot learn this other than from watching it play out," he says. He is planning experiments — some of which might stretch past day 14 — to try to find out exactly how an embryo containing some aneuploid cells adjusts.

Beyond two weeks

Researchers working with animal embryos have already cultured them beyond the developmental equivalent of 14 days, which could pave the way for similar advances in human embryo culture. In March, Jacob Hanna's group at the Weizmann Institute of Science in Rehovot, Israel, doubled the time that mouse embryos can be cultured in the laboratory⁷, from day 5.5 to day 11 (roughly equivalent to human days 13 to 30). Although other researchers describe the process as finicky, the embryos go well into the process of developing organs.

Both Brivanlou and Zernicka-Goetz plan to push the envelope for culturing human embryos further, doing for humans what Hanna's team has done for mice. Brivanlou wants to unlock the genetic program that turns stem cells into the first brain cells, and also wants to reveal the molecular instructions for a four-chambered beating heart. Both events arise after day 14 and unravelling them could prove key to understanding neurodevelopmental disorders and common congenital heart defects.

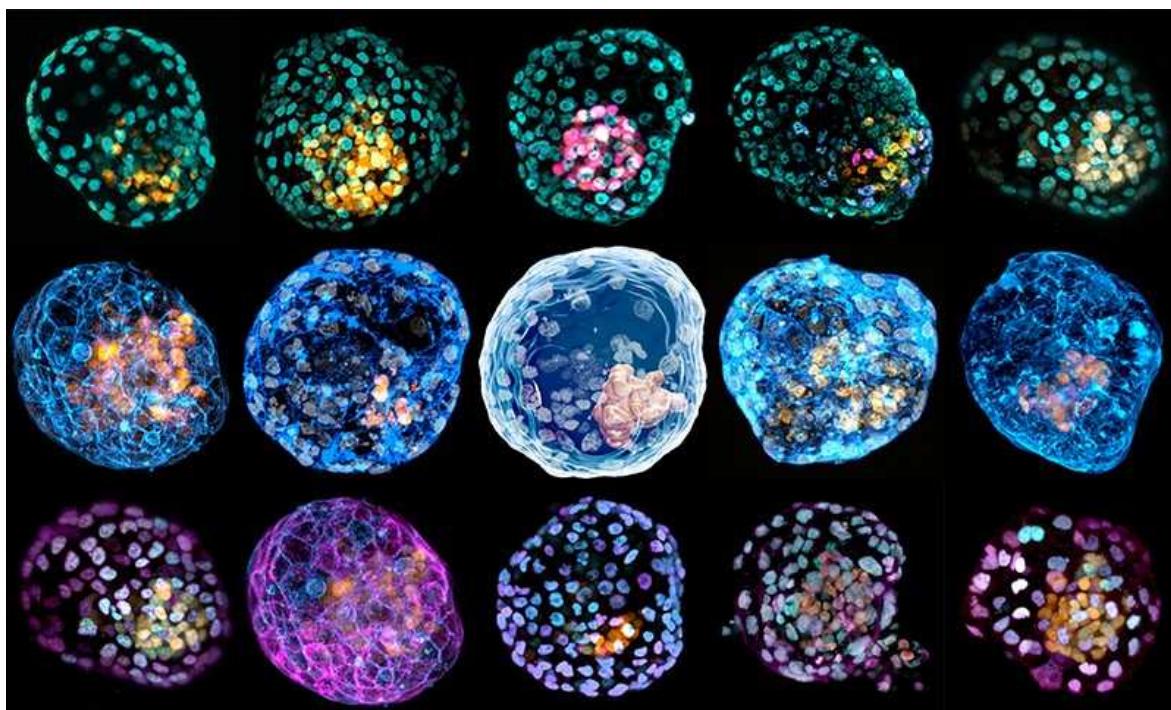
Brivanlou and others say that, in the future, researchers will be technically capable of watching a human embryo implant into uterine tissue in the lab.

His group is applying to his university's regulatory committee to extend their research on human embryos to day 21. In that extra week, the entire body plan is set, including the structures that will become spinal cord, brain, bone, heart, blood, muscles and face. "I don't know if we can push a human embryo past day 14," he says. "But once we get there, we should move very gently."

Model embryos

Researchers already have some ways of looking at the process of gastrulation in mammals, not by using real embryos, but by constructing models of them from 3D mixtures of stem cells.

In the past five years, researchers have crafted a variety of embryo models in the lab that can be used to glimpse stages beyond day 14. In most cases, these embryo models are not subject to the 14-day rule or any special review.

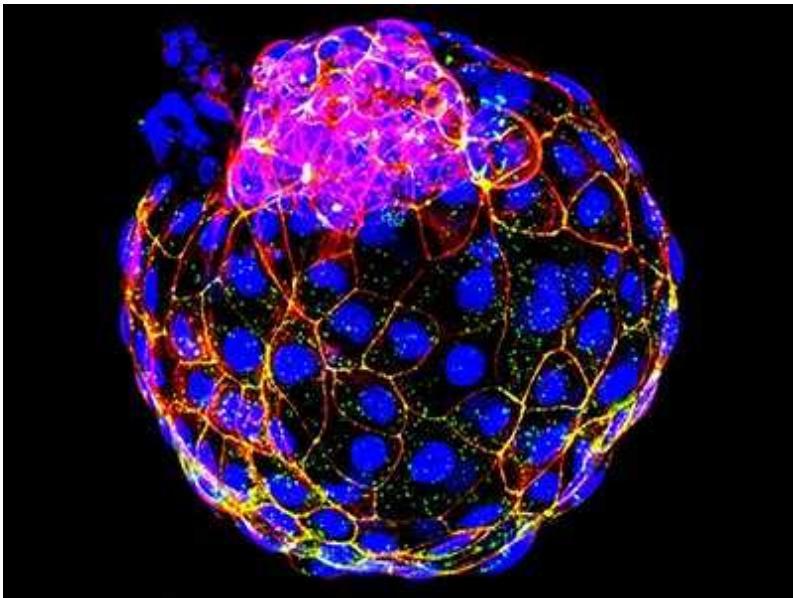


These embryo models, which mimic the blastocyst stage, were reprogrammed from skin cells. Credit: Monash University

In 2017, Jianping Fu, a biomedical engineer at University of Michigan in Ann Arbor, and his team made the first human embryo model simply by putting ES cells into 3D culture, where they self-organized to form the amniotic sac and the first signs of a primitive streak⁸. The finding generated a great deal of excitement in the field, Fu says. “That such possibilities exist within human ES cells — they’re amazing.”

Models like Fu’s proliferated, and can now mimic portions of the earliest embryonic stages in mouse and human — implantation⁸, gastrulation^{9,10}, and the rudimentary beginnings of brain¹¹, spinal cord¹² and heart development. For instance, Fu and others have created human neuruloids¹², which model the formation of precursors to spinal cord and brain, and which Fu says will help researchers culture functional neurons that could be placed into patients one day. The mixtures of cells often look and behave like embryos at the same stage, but whether they recapitulate the molecular and cellular events of normal development remains a yawning gap in knowledge.

Fu and others acknowledge that, as human embryo models form more complex structures and push further along the developmental timeline, they raise new ethical questions. For instance, their neurons could begin to fire or their heart cells to beat. Or they could acquire the potential to develop beyond the limited stages they mimic so far. Most models lack the full complement of life-supporting tissues that they would need to form a whole embryo, but in the past few years a handful of labs have constructed mouse and human models of the blastocyst stage, called ‘blastoids’. These contain the precursors to these support tissues and could theoretically form the entire organism^{13–15}.



Lab-grown structures mimic human embryo's earliest stage yet

The ISSCR has its eye on the area, and its guidelines state that models that contain these supporting tissues must be subject to special oversight and grown for the minimum time necessary to meet the scientific aim. As they evolve, these models will need ethical reconsideration, says Robin Lovell-Badge, a stem-cell biologist at the Francis Crick Institute who chaired the ISSCR steering committee. “Obviously, this space has to be watched.”

As work in real and model embryos moves forward, scientists are keen to know how similar the two really are. Finding out how models differ in their molecular details, and how their cells behave, is the main reason researchers wish to push beyond 14 days in real embryos. “We can learn a lot from a model,” says Jesse Veenvliet, a developmental biologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. “But it’s important to know where it goes wrong.”

In 2020, Moris and colleagues ran a side-by-side gene expression comparison of mouse embryos and mouse gastruloids and found striking similarities in the timed waves of signalling that set up the body plan¹⁶. She says that this exact kind of benchmarking should be done for human gastruloids, too, which would require culturing human embryos up to around day 21. Rivron envisages molecular maps for each stage of human development.

Once enough benchmarking is done, researchers could reduce the number of human embryos used, only turning to them when there is a strong justification.

Stop the clock

But for just how long should researchers watch human development play out in a dish? Critics have said it was irresponsible for the ISSCR to relax the 14-day rule without giving researchers a new stopping point — that it gives the appearance of a green light for embryo research.

Fu was on the ISSCR committee that developed the new guidelines over 18 months and in more than 100 Zoom meetings. “The science is progressing so fast, it was hard to draw another stop sign as a scientific community,” he says. Instead, the guidelines left the door open to do research as long as the appropriate review processes occur.

Brivanlou concurs with the decision. “In my conscience, I know there are great benefits in pushing past 14 days,” he says, pointing to the work on abnormal chromosomes. “It may literally save lives in the next generation.” He and Lovell-Badge argue that it would be unethical not to allow some post-14-day research because it could unlock how organ cell types arise and how miscarriages and birth defects occur.

It’s also becoming fuzzier as to which experiments and models should be accorded the same status as post-14-day embryo research. The ISSCR guidelines draw a clear ethical division, placing only models that include supporting tissues — and that theoretically have the potential to develop fully — in the same category as post-14-day embryos. (The guidelines also ban the transfer of human research embryos, human-animal chimaeric embryos or human embryo models into an animal or human uterus.)

Some researchers are conservative about the lifting of the 14-day line. Moris, for instance, doesn’t think that the public has had a real chance to weigh the consequences of post-14-day experimentation.

Bioethicist Josephine Johnston goes one step further: “I think it’s a mistake to drop the 14-day rule and not propose another rule.” A limit signals that

the scientific community understands that society values human embryos and respects that, says Johnston, a bioethicist at the Hastings Center in Garrison, New York. Dropping the limit “has the potential to really shake public trust”. In addition, researchers need to better explain how studying embryos beyond 14 days “would actually help humanity”, she says. They also need to be upfront about the details of research on embryos, Johnston says. “A lot of this research feels very distant, but it’s careless not to have some limits.”

Nature **597**, 22–24 (2021)

doi: <https://doi.org/10.1038/d41586-021-02343-7>

References

1. 1.

Lovell-Badge, R. *et al.* *Stem Cell Rep.* **16**, 1398–1408 (2021).

2. 2.

Shahbazi, M. N. *et al.* *Nature Cell Biol.* **18**, 700–708 (2016).

3. 3.

Deglincerti, A. *et al.* *Nature* **533**, 251–254 (2016).

4. 4.

Molè, M. A. *et al.* *Nature Commun.* **12**, 3679 (2021).

5. 5.

Shahbazi, M. N. *et al.* *Nature Commun.* **11**, 3987 (2020).

6. 6.

Yang, M. *et al.* *Nature Cell Biol.* **23**, 314–321 (2021).

7. 7.

Aguilera-Castrejon, A. *et al.* *Nature* **593**, 119–124 (2021).

8. 8.

Shao, Y. *et al.* *Nature Commun.* **8**, 208 (2017).

9. 9.

Sozen, B. *et al.* *Nature Cell Biol.* **20**, 979–989 (2018).

10. 10.

Moris, N. *et al.* *Nature* **582**, 410–414 (2020).

11. 11.

Xue, X., Wang, R. P. & Fu, J. *Curr. Opin. Biomed. Eng.* **13**, 127–133 (2020).

12. 12.

Veenvliet, J. V. *et al.* *Science* **370**, eaba4937 (2020).

13. 13.

Rivron, N. C. *et al.* *Nature* **557**, 106–111 (2018).

14. 14.

Yu, L. *et al.* *Nature* **591**, 620–626 (2021).

15. 15.

Liu, X. *et al.* *Nature* **591**, 627–632 (2021).

16. 16.

van den Brink, S. C. *et al.* *Nature* **582**, 405–409 (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02343-7>

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

Books & Arts

- **Witness in US climate-change law suit tells all** [30 August 2021]
Book Review • Was the government a primary cause of the US addiction to fossil fuel? An expert who testified gives evidence.
- **Shells, sweat, and the fear of strangers: Books in brief** [24 August 2021]
Book Review • Andrew Robinson reviews five of the week's best science picks.

- BOOK REVIEW
- 30 August 2021

Witness in US climate-change law suit tells all

Was the government a primary cause of the US addiction to fossil fuel? An expert who testified gives evidence.

- [Catherine Higham](#) ⁰



Young people rally outside the US Supreme Court in 2018 to support the climate case *Juliana et al. v. US government*. Credit: Win McNamee/Getty

They Knew: The US Federal Government's Fifty-Year Role in Causing the Climate Crisis James Gustave Speth MIT Press (2021)

The week of 24 May 2021 was, in the words of one leading litigator, “a great week to be a climate lawyer”. Two landmark judgments sent shock waves through the legal world. On 26 May, the Hague District Court ordered Royal Dutch Shell to reduce its greenhouse-gas emissions by 45% from 2019 levels by 2030. It was the first time that a court had ordered a company to take such action on the basis of harm to others. The next day, an Australian judge found that the minister of the environment had a duty of care to Australian children to protect them from foreseeable future harms caused by increased greenhouse-gas emissions in the atmosphere.

Both judgments represent years of graft by committed plaintiffs and legal teams. They are also part of a wider story. Around the world, some 1,850 ‘climate-change cases’ have now been filed, many by climate lawyers and clients who have turned to the courts to address the climate crisis by filling the gaps left by slow or inadequate government action. Successes such as those in Australia and the Netherlands, unimaginable a decade ago, represent hard-fought victories in a continuing battle over the evolving understanding of rights and responsibilities in the context of a changing climate. Neither would have been possible without previous cases — won or lost — that introduced climate science into the courtroom.



Why I broke the law for climate change

It is against this backdrop that Gus Speth's timely book must be understood. It began as an expert report in a famous climate-change case: *Juliana et al. v. US Government*, submitted to the Oregon District Court by 21 youth plaintiffs and others in 2015. The claimants' arguments rested on the federal government's constitutional obligations to protect US citizens' rights to life, liberty and property, and to protect the atmosphere as part of a "public trust". They alleged that the government had violated these duties by "perpetuating" a fossil-fuel-based energy system, despite "long-standing knowledge" of its harms.

In their introduction to the book, Julia Olson and Philip Gregory, counsel for the plaintiffs, explain that they had to demonstrate that the government had both allowed and knowingly created the alleged harms. This required them to prove that the conduct of the federal government was a primary cause of the US addiction to fossil fuels, the point Speth's report was commissioned to demonstrate.

Speth was an inspired choice of witness. Drawing on first-hand knowledge from his time as chair of the US Council on Environmental Quality during the administration of president Jimmy Carter (1977–81), and his founding role in several major environmental non-governmental organizations, Speth gives a clear and concise account of the scientific evidence available to successive US presidents and Congresses over five decades. He provides a chilling description of the gulf between the safer course of action recommended by scientists and advisers, and the reality of federal policy. Given the book's original purpose — an exhaustive presentation of evidence — at times it can seem like a laundry list of warnings ignored. Fortunately, Speth peppers it with gripping insider details.



[Not slashing emissions? See you in court](#)

One compelling example: the speed with which the administration of George H. W. Bush (1989–93) turned away from early recognition of the benefits of using policy to reduce climate risk, and became dominated by staffers who used material from the fossil-fuel industry to cast doubt on the science. The administration of Bill Clinton (1993–2001) ran a miscalculated public-awareness campaign that actually increased the partisan divide in perceptions of climate science. These episodes offer fascinating insight into subsequent events.

Much of the story is not new. Elements of the interactions between the US fossil-fuel industry and the federal government have been charted elsewhere, such as in the work of science historian Naomi Oreskes, or in climate scientist Michael Mann's book *The New Climate War* (2021). Yet even for the familiar reader, Speth's focus on the federal government provides a fresh perspective.

A key feature of the Australian and Dutch cases is that, unlike many before them, they establish a legal responsibility for the defendants to act now and in the future. The important issue was not what the minister or Shell knew in the past, but what they should do now, in the light of the global consensus on the climate threat. Although establishing the parameters for future decisions is a crucial goal of climate litigation, Speth's book reminds us that major

questions about historic responsibility cannot be ignored. In the US context, the *Juliana* plaintiffs have begun settlement talks with the administration of President Joe Biden and vice-president Kamala Harris, which might lead to answers. Around the world, similar reckonings with the past are just beginning.

Nature **597**, 25-26 (2021)

doi: <https://doi.org/10.1038/d41586-021-02344-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02344-6>

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

- BOOK REVIEW
- 24 August 2021

Shells, sweat, and the fear of strangers: Books in brief

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

- [Andrew Robinson](#) 0

THE SOUND *of* THE SEA

SEASHELLS *and the*
FATE *of the* OCEANS

CYNTHIA BARNETT



The Sound of the Sea

Cynthia Barnett *W. W. Norton* (2021)

“Seashells were money before coin, jewellery before gems, art before canvas,” says science writer Cynthia Barnett in her arresting meditation on shells and ocean history. Consider *Monetaria moneta*, the money cowrie. Carried as ships’ ballast, it was the main currency used in West Africa for the trans-Atlantic trade in enslaved people. Oil company Shell — subject of one chapter — grew from a nineteenth-century firm selling East Asian seashells for interior decor in Europe. Its ship *Murex*, named after a spiny mollusc, was the first oil tanker to traverse the Suez Canal.

FLASHES OF CREATION

GEORGE GAMOW, FRED HOYLE,

AND THE GREAT
BIG BANG DEBATE

PAUL HALPERN

Flashes of Creation

Paul Halpern *Basic* (2021)

Astrophysicist Fred Hoyle coined the term Big Bang in a 1949 radio broadcast to mock the theory of cosmologists including George Gamow, who later popularized the idea in his 1952 book *The Creation of the Universe*. Hoyle preferred a never-ending process of creation, a ‘steady state’. Albert Einstein also distrusted the idea. Physicist Paul Halpern retells this conflict in his highly engaging history, while noting that today’s dominant Big Bang cosmology leaves dark energy and cold dark matter unexplained.

THE JOY OF SWEAT

**F
E
E
THE STRANGE
SCIENCE OF
PERSPIRATION**

SARAH EVERTS

The Joy of Sweat

Sarah Everts *W. W. Norton* (2021)

Most animals do not sweat to regulate their body temperature. Some evolutionary biologists even argue that perspiration helped humans to dominance, notes science journalist Sarah Everts in her well-researched, zesty study. Certainly, we are ambivalent about it. Most cultures institutionalize sweating, for example in gyms, saunas and bathhouses. Yet we also spend significant money on deodorants and laundry products. “Sweat may be sticky, stinky and gross, but it’s also one of our most fascinating and little understood secretions.”

"Utterly original and
beautifully written."

ZIA HAIDER RAHMAN

OF FEAR
AND STRANGERS
A HISTORY OF
XENOPHOBIA

GEORGE
MAKARI

Of Fear and Strangers

George Makari *Yale Univ. Press* (2021)

Historian and psychiatrist George Makari comes from a Lebanese family that emigrated to the United States, he explains in a prologue to his diverse and scholarly history of xenophobia, provoked by the US election of Donald Trump in 2016 and the UK Brexit referendum. “Xenophobia had come back from the dead,” he says, and was no longer “some antiquated, classical term”. He discovers that despite its ancient Greek etymology — *xenos* for ‘stranger’ — the term dates from the late nineteenth century: the age of imperialism.

BERNOULLI'S FALLACY

Statistical Illogic
and the Crisis of
Modern Science

AUBREY CLAYTON

Bernoulli's Fallacy

Aubrey Clayton *Columbia Univ. Press* (2021)

The COVID-19 pandemic reminds us of the importance of statistics and the difficulty of interpreting them. “Probability opens doors for us and then trips us up,” comments mathematician Aubrey Clayton in his densely challenging history. He rebuts the argument of seventeenth-century mathematician Jacob Bernoulli that data alone permit objective inference, which ignores the subjectivity inherent in assigning probability. Scientists must reject the authority of frequentist statistics, he argues: better to adopt a Bayesian approach.

Nature **597**, 26 (2021)

doi: <https://doi.org/10.1038/d41586-021-02308-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02308-w>

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

Opinion

- **[Carolyn S. Shoemaker \(1929–2021\)](#)** [26 August 2021]
Obituary • Co-discoverer of first comet known to collide with a planet in modern times.
- **[Food systems: seven priorities to end hunger and protect the planet](#)** [30 August 2021]
Comment • Here's how the United Nations should harness science and technology to improve nutrition and safeguard the environment.
- **[Early-career researchers: choose change, not complicity](#)**
[31 August 2021]
Correspondence •
- **[Boost for Africa's research must protect its biodiversity](#)** [31 August 2021]
Correspondence •
- **[In memory of a game-changing haematologist](#)** [31 August 2021]
Correspondence •
- **[Australian bush fires and fuel loads](#)** [31 August 2021]
Correspondence •

- OBITUARY
- 26 August 2021

Carolyn S. Shoemaker (1929–2021)

Co-discoverer of first comet known to collide with a planet in modern times.

- [David H. Levy](#) 0



Credit: University of Chicago Library

For many years, Carolyn S. Shoemaker held the record for the largest number of comets discovered by an individual, but by far her most famous discovery was comet Shoemaker–Levy 9. From 16 to 22 July 1994, fragments of this comet, travelling at some 60 kilometres per second, collided with Jupiter, resulting in the most dramatic explosions in the Solar System ever witnessed by humanity. The dark spots left by the impacts were visible for almost a year. This singular experience had begun almost 16 months previously at the Palomar Observatory, California, when Shoemaker stopped scanning her photographic plates, looked up and said, “I do not know what I have, but it looks like a squashed comet.” In the following few minutes, her husband Gene and I confirmed the sighting.

Shoemaker’s achievements went far beyond this discovery. Between 1980 and 1994, as a member of the Palomar Asteroid and Comet Survey (PACS), she found 32 comets, plus more than 400 asteroids. Although PACS’s objective was to find asteroids or comets that could pose a threat to civilization, the discovery of Shoemaker–Levy 9 completely overshadowed that aim. The interest generated by the comet’s impacts with Jupiter was almost as spectacular as the collisions themselves. For the first time, people the world over grappled with questions about what transpires when comets strike planets, and how these impacts might offer an insight into the origins of life on Earth.

Born in Gallup, New Mexico, Shoemaker grew up in Chico, California. At what is now California State University, Chico, she earned bachelor’s and master’s degrees in history, political science and English literature. She had no interest in science. However, at her brother’s wedding, she met geologist Gene Shoemaker, who was also her brother’s roommate at the California Institute of Technology in Pasadena. Carolyn was enthralled by how Gene brought geology to life, and they married in 1951. She left her career as a schoolteacher to begin raising their three children, while her husband trained astronauts to conduct geology experiments during moonwalks.



Eugene M. Shoemaker (1928-97)

As the children grew older, he suggested that she join his new programme to search for near-Earth asteroids. She proved to have a keen eye and an uncanny ability to use the stereomicroscope to look for differences in photographs of the same patch of sky taken short intervals apart, discovering her first comet on 3 September 1983. By the time I joined the team in 1989, Carolyn had found 17 new comets. During this time, she became familiar with the idiosyncrasies of the first and oldest telescope at Palomar, the 46-centimetre ‘Schmidt camera’. It was never converted to today’s electronic standards; instead, it used only photographic film. The most successful film Carolyn and Gene used was a very fine-grained Kodak 4415, hypersensitized in an oven for 6 hours beforehand.

Each observation began by loading a 15-centimetre disc of film into a placeholder, mounting it into the telescope, and exposing the film for 8 minutes. It would then be placed in a light-tight box, and a new film inserted. Our nights would be divided into sets of four, sometimes five, fields; we then repeated each exposure. At our most efficient, the time between the end of one exposure and the start of the next was as little as 90 seconds. Our nights would be divided into observations of four, sometimes five, fields of the sky; we then repeated each exposure. It was during one of these routine sessions that we recorded the two ‘discovery’ images of

Shoemaker-Levy 9. When Brian G. Marsden at the Harvard-Smithsonian Center for Astrophysics calculated that the comet was on a collision course with Jupiter, Gene thought: “In my lifetime, I am going to witness a cosmic impact.” Carolyn thought: “I am going to lose one of my comets.”

One consequence of these observations is that, since 1994, cosmic impacts have been taken seriously, and more programmes around the world are looking for comets and asteroids that could pose a threat to Earth. Other lines of research are pursuing the idea that life on Earth might have been seeded by simple organic molecules arriving from space on comets. And, parenthetically, the ‘giggle factor’ — the offhand reaction of journalists and laypeople to the idea of objects from space hitting Earth — has dissipated completely.

After the impacts, Carolyn resumed her search with PACS, along with her husband and me. Although the programme concluded at the end of December 1994, the team continued the work with two smaller Schmidt cameras at the Jarnac Observatory in Arizona. Gene was killed in a car accident in Australia in 1997. Carolyn bravely continued her work after that.

Over the years at Palomar, there were times when Carolyn needed to remind the rest of her team why our work was so important. During a rare display of the Northern Lights at Palomar, she forced Gene to stop observing for a short while and enjoy the shimmer in the sky. She understood clearly that although the scientific goals mattered, PACS would not have been worthwhile had the team members not retained their sense of wonder.

Carolyn never really considered herself a scientist, although she absorbed a large amount of geology and astronomy from her husband. She was a forthright person, gifted with inordinate patience and a fine sense of humour. Before the Jupiter impacts, a reporter asked her what would happen if all the comet’s fragments were to hit Earth instead. “We would all die,” she answered. The interviewer explained that this was for a children’s programme, then posed the question again. Carolyn’s second answer: “We would all be very uncomfortable.”

doi: <https://doi.org/10.1038/d41586-021-02345-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02345-5>

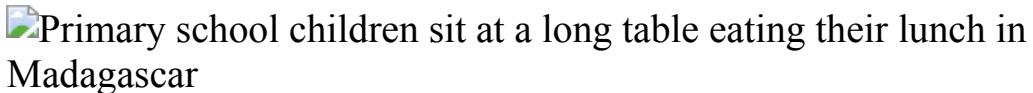
| [Section menu](#) | [Main menu](#) |

- COMMENT
- 30 August 2021

Food systems: seven priorities to end hunger and protect the planet

Here's how the United Nations should harness science and technology to improve nutrition and safeguard the environment.

- [Joachim von Braun](#) ⁰,
- [Kaosar Afsana](#) ¹,
- [Louise O. Fresco](#) ² &
- [Mohamed Hassan](#) ³



School children in Madagascar eat lunch provided as part of a nutrition initiative run by the World Food Programme. Credit: Rijasolo/AFP/Getty

The world's food system is in disarray. One in ten people is undernourished. One in four is overweight. More than one-third of the world's population cannot afford a healthy diet. Food supplies are disrupted by heatwaves, floods, droughts and wars. The number of people going hungry in 2020 was 15% higher than in 2019, owing to the COVID-19 pandemic and armed conflicts¹.

Our planetary habitat suffers, too. The food sector emits about 30% of the world's greenhouse gases. Expanding cropland, pastures and tree plantations drive two-thirds of the loss in forests (5.5 million hectares per year), mostly in the tropics². Poor farming practices degrade soils, pollute and deplete water supplies and lower biodiversity.

As these interlinkages become clear, approaches to food are shifting — away from production, consumption and value chains towards safety, networks and complexity. Recent crises around global warming and COVID-19 have compounded concerns. Policymakers have taken note.



Counting the hidden \$12-trillion cost of a broken food system

In September, the United Nations secretary-general, António Guterres, will convene a Food Systems Summit. This is only the sixth UN summit on food since 1943, and the first with heads of states in the UN General Assembly. A group of leading scientists, known as the Scientific Group, has been tasked with ensuring that the science underpinning the 2021 summit is robust, broad and independent. We, the authors, are this group's chair and vice-chairs. Although such approaches are familiar in areas such as climate change and biodiversity, this marks the first time that scientists have been explicitly brought in to multilateral discussions around food (see [*Nature* 595, 332; 2021](#)).

The global food system needs a revamp — in policies and institutions as well as on social, business and technology fronts³. Science is one lens for making sure that changes are integrated and collectively deliver better outcomes. But the task is challenging. Food spans many disciplines — not least agriculture, health, climate science, artificial intelligence and digital science, political science and economics. The indirect, adverse effects of

policies on climate change, biodiversity loss and health need to be factored in to the true costs of food; these are estimated to be about twice the current market value of food consumption globally⁴. A range of voices is crucial. The Scientific Group is engaging with hundreds of experts across civil society, including Indigenous peoples, producer and youth organizations and the private sectors.

Here we highlight the key roles that scientists should take to accelerate the transformation to healthier, more sustainable, equitable and resilient food systems. These seven priorities reflect the Scientific Group's evidence base, comprising more than 50 reports and briefs (see go.nature.com/3dtoazu).



A farmer in Bolivia harvests quinoa for its edible seeds. Credit: Giraudou Laurent/Sygma/Getty

Seven priorities

Science-driven advances are needed to address the following challenges.

End hunger and improve diets. Scientists need to identify optimal conditions and opportunities for investments to make healthy and nutritious foods more available, affordable and accessible. Measures that jointly improve more than one of these are most effective. For example, increased irrigation on small farms in Tanzania and Ethiopia has enhanced productivity, dietary diversity and farmers' incomes⁵.

Three big game-changers are: enhancing research and development in agriculture and food to increase productivity in a sustainable way; slashing food waste and losses; and adding income and nutrition components to social-protection programmes. Research priorities to cut waste include scaling up solar energy and battery storage technologies to make food processing and preservation more affordable. New forms of packaging using recycled materials, coatings of nanomaterials and even edible films would keep foods fresh for longer. School feeding programmes, together with incentives to keep children in education (such as take-home rations for parents) have seen success in Mali, for example, where they increased school enrolment by 10 percentage points⁶. Under COVID-19 lockdowns, these types of programme became even more relevant.



Europe's Green Deal offshores environmental damage to other nations

Researchers also need to study behavioural barriers to healthy eating, such as snacking under stress. They should develop policy guidelines for

educational food labels, and model the impacts of putting taxes and regulations on unhealthy foods (such as sugar and trans-fats). The health properties of fortified foods and cultivated meats must also be established.

De-risk food systems. The more global, dynamic and complex food systems become, the more open they are to new risks. Scientists need to improve how they understand, monitor, analyse and communicate such vulnerabilities. For example, droughts, the expansion of biofuels and financial speculation after the sudden imposition of trade barriers led to food price hikes in 2008⁷. The COVID-19 pandemic and armed conflicts have shaken food value chains across Africa this year, driving up food prices. Successful initiatives exist, combining on-the-ground observations of food systems and nutrition with forecasting. These include FEWS NET (<https://fews.net>) and the joint analyses from the UN Food and Agriculture Organization and the World Food Programme on early warnings of food insecurity⁸.

Policies and economic solutions are needed. For example, new insurance products aided by remote sensing and weather forecasts would provide cover for lost crops and livestock. Solar-powered irrigation systems would reduce risk from drought. Smartphone apps would provide farmers with information on local crop pests, weather risks and market opportunities; these are already used in Kenya, Senegal, India and Bangladesh⁹. Payment schemes are needed to encourage farmers to manage and capture carbon in soils and trees, and to trade it.



A seaweed farmer in Bali, Indonesia.Credit: Anton Raharjo/NurPhoto/Getty

Protect equality and rights. Poverty and inequalities associated with gender, ethnicity and age restrict many people's access to healthy foods. Socio-economic researchers need to suggest inclusive ways to transform the more than 400 million smallholder farms worldwide. They must identify pathways out of inequitable and unfair arrangements over land, credit and labour, and empower the rights of women and youth. For example, if female-headed households in southern Ethiopia had the same resources as male-headed ones, their productivity in maize (corn) would increase by more than 40%, to match that of the latter¹⁰.

Protecting the land rights of smallholders, women and Indigenous peoples is paramount. Technology can ensure transparency and efficiency. For example, using blockchain ledgers of ownership rights to allocate land could be an opportunity in Ghana¹¹. At the trans-national scale, [the Land Matrix Initiative](#) collects and shares data on big land acquisitions and investments in low- and middle-income countries; it covers deals in almost 100 nations worldwide. Similar solutions are needed to protect the land rights of

Indigenous peoples¹². Also required are efforts to build local research capacity, educational programmes around food and farming, and training and financing opportunities in rural areas.

Boost bioscience. Researchers need to find ways to restore soil health and improve the efficiency of cropping, crop breeding and recarbonizing the soil and biosphere. Linkages among all Earth systems must be considered together — known as a One Health approach (see go.nature.com/3jy7ekh).



Five priorities for a sustainable ocean economy

Alternative sources of healthy protein need to be advanced, such as plant-based and insect-derived proteins, including for animal feed. Plant-breeding techniques that capture nitrogen from the air, to reduce the need for fertilizers and increase nutrients, should be investigated. Genetic engineering and biotechnology should be applied to increase the productivity, quality and resistance of crops to pests and drought. Recent examples include banana varieties that are resistant to Fusarium wilt fungal diseases, and pest-resistant *Bt* aubergines. To widen access to bioscience technologies, intellectual-property rights, skills and data sharing should be addressed.

Protect resources. Tools are needed to help people to manage soils, land and water sustainably. For example, hand-held digital devices and remote sensing can track concentrations of soil carbon and other nutrients.

Artificial-intelligence systems and drones allow farmers to spot areas that need irrigation, fertilization and protection from pests. Soil microbes can be harnessed to improve soil structure, carbon storage and yields. Researchers need to adapt and scale up such technologies.

Biodiversity and genetic bases need to be protected. Seed varieties must be preserved, and their phenotypes and genotypes explored in the contexts of climate change and nutrition. Traditional food and forest systems, including those of Indigenous peoples, need to be better understood and supported in national agricultural research systems. Cooperation for mutual benefit should be explored, as has been done for climate adaptation in US Indigenous areas^{[13](#)}.



Averting hunger in sub-Saharan Africa requires data and synthesis

Sustain aquatic foods. Most of the focus on food so far has been on soil-based agriculture. Fish, shellfish and aquatic plants such as seaweed have much to offer nutritionally and environmentally. [Aquatic foods need to be better integrated](#) into the understanding of food systems^{[14](#)}. Researchers should look for ways to increase nutritional diversity in aquatic foods and sequester carbon in marine and freshwater environments.

Ecological-science perspectives and global cooperation and institutions are needed to make harvesting of oceans, coastal waters and freshwater

resources sustainable and protect biodiversity. The sustainability of fish-feeding systems needs attention; for example, by exploring the use of insect rearing, oil-rich modified legumes and micro-algae as fish feed.

Harness digital technology. Robots, sensors and artificial intelligence are increasingly used on farms: to harvest crops and milk cows, for example. Sensors can monitor the origin and quality of ingredients and products along the food-processing chain to reduce losses and guarantee food safety. But most farmers and producers still don't have access. To spread the benefits, devices need to become cheaper and easier to purchase and use. Rental services similar to Uber for farm machinery should be developed, as has been done with tractors in India. Rural electricity supplies will have to be expanded, along with IT training and education.

First steps

The 2021 Food Systems Summit is a great opportunity to end hunger by 2030 and set in train a sustainable food system. Previous summits have delivered change: creating the FAO (after 1943); strengthening the global food-research partnership CGIAR and founding the International Food Policy Research Institute (after 1974); accelerating the human right to food (2002); and establishing monitoring systems to warn of food-price crises (2009).



Without food, there can be no exit from the pandemic

The breadth of the 2021 agenda could be a hindrance, however, to achieving its goals. To avoid failure, delegates should focus. They should prioritize establishing a guiding framework — for transforming diverse national and local food systems, as well as global networks, with the challenges of trade, finance, climate, innovation and governance.

Debates will be fierce. Food is a contentious topic. Disagreements abound over goals, pathways and speed of change, and the roles of science and technology, the private sector and the UN. For example, some see agroecology as the only acceptable way of farming; some view biotechnology and gene editing as dangers; others see opportunities. The Scientific Group has aimed to offer a scientific basis to this diversity of perspectives.

Actions and targets

Once plans are agreed, the UN Food Systems Summit will need to move to implementation. Here are our suggestions.



Nature-based solutions can help cool the planet — if we act now

First, boost finance. On the research front, we propose that governments allocate to food research at least 1% of the proportion of their nations' gross domestic product that relates to food systems. Many countries spend only half of that. Least-developed countries should be given aid to reach a similar level. To end hunger for the poorest, we propose the establishment of a special fund. This would be supported by development-aid donors and bonds backed by the International Monetary Fund and World Bank. Research and modelling would be required on implementation and impacts.

Second, increase scientific capacity. The special fund could be used to strengthen research capacity in low- and middle-income countries, and expand research collaborations between the public and private sectors, and among farmers, start-up firms in food value chains and science communities. Sharing research infrastructure and data between the global south and global north would be a good start.

Third, strengthen science–policy interfaces. In stark contrast to many other fields, agriculture, food security and nutrition do not have an international agreement or convention to consolidate actions. We call on the UN Food Systems Summit and UN member states to explore an intergovernmental treaty or framework convention on food systems, analogous to the conventions on climate, biodiversity and desertification agreed on in Rio de Janeiro in 1992. The framework will need to include a strong independent scientific body that provides policy advice in the follow-up to the summit. We recommend that all science organizations and academies with food-relevant research be included in a preparatory process.

Bringing the tools of science to the table will help to transform the global food system to end hunger and achieve the UN Sustainable Development Goals by 2030.

Nature **597**, 28–30 (2021)

doi: <https://doi.org/10.1038/d41586-021-02331-x>

References

1. 1.

Food & Agriculture Organization, International Fund for Agricultural Development, UNICEF, World Food Programme and World Health Organization. *The State of Food Security and Nutrition in the World 2021* (FAO, 2021).

2. 2.

Pendrill, F., Persson, U. M., Godar, J. & Kastner, T. *Environ. Res. Lett.* **14**, 055003 (2019).

3. 3.

Organisation for Economic Co-operation and Development. *Making Better Policies for Food Systems* (OECD, 2021).

4. 4.

Hendriks, S. et al. *The True Cost and True Price of Food* (United Nations, 2021).

5. 5.

Passarelli, S., Mekonnen, D., Bryan, E. & Ringler, C. *Food Sec.* **10**, 981–997 (2018).

6. 6.

Aurino, E., Tranchant, J.-P, Diallo, A. S. & Gelli, A. *J. Dev. Stud.* **55** (Suppl. 1), 7–28 (2019).

7. 7.

Kalkuhl, M., von Braun, J. & Torero, M. (eds) *Food Price Volatility and Its Implications for Food Security and Policy* (Springer, 2016).

8. 8.

Food & Agriculture Organization & World Food Programme. *FAO-WFP Early Warning Analysis of Acute Food Insecurity Hotspots, July*

2020 (FAO & WFP, 2020).

9. 9.

Baumüller, H. In *Handbook on ICT in Developing Countries: 5G Perspectives* (eds Skouby, K. E., Williams, I. & Gyamfi, A.) 191–210 (River, 2017).

10. 10.

Gebre, G. G., Isoda, H., Rahut, D. B., Amekawa, Y. & Nomura, H. *GeoJournal* **86**, 843–864 (2021).

11. 11.

Mintah, K., Boateng, F. G., Baako, K. T., Gaisie, E. & Otchere, G. K. *Land Use Policy* **109**, 105635 (2021).

12. 12.

Food & Agriculture Organization. *The White/Wiphala Paper on Indigenous Peoples' Food Systems* (FAO, 2021).

13. 13.

Tribal Adaptation Menu Team. *Dibagin jigaadeg Anishinaabe Ezhitwaad: A Tribal Climate Adaptation Menu* (Great Lakes Indian Fish and Wildlife Commission, 2019).

14. 14.

Lubchenco, J., Haugan, P. M. & Pangestu, M. E. *Nature* **588**, 30–32 (2020).

- CORRESPONDENCE
- 31 August 2021

Early-career researchers: choose change, not complicity

- [Lucrecia K. Aguilar](#) 0

Early-career researchers generally are ardent supporters of greater diversity, equity and inclusion, work–life balance and mental well-being in academia. Yet the precariousness of our careers seems to demand a default to an academic system that perpetuates injustices and poor quality of life ([K. N. Laland *Nature* 584, 653–654 \(2020\)](#); [E. N. Satinsky *et al.* *Sci. Rep.* 11, 14370; 2021](#)). We must apply the changes we wish to see in academia to our own lives if the system is to work better for everyone.

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.

Additional access options:

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

Nature **597**, 31 (2021)

doi: <https://doi.org/10.1038/d41586-021-02355-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02355-3>

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

- CORRESPONDENCE
- 31 August 2021

Boost for Africa's research must protect its biodiversity

- [Nils Chr. Stenseth](#) ORCID: <http://orcid.org/0000-0002-1591-5399>⁰
&
- [Sebsebe Demissew](#) ORCID: <http://orcid.org/0000-0002-0123-9596>¹

We write on behalf of 209 scientists (see go.nature.com/3sa16p9) to endorse a new initiative by the African Research Universities Alliance and the Guild of European Research-Intensive Universities (see go.nature.com/3b364hj). This calls for greater investment by the African Union and the European Union in Africa's universities, to help them address global challenges such as public health, climate change and good governance. We strongly encourage expansion of the initiative to encompass environmental and biodiversity issues that are crucial to the continent's future.

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.

Additional access options:

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

Nature **597**, 31 (2021)

doi: <https://doi.org/10.1038/d41586-021-02356-2>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02356-2>

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

- CORRESPONDENCE
- 31 August 2021

In memory of a game-changing haematologist

- [Daniel Lucas](#) ⁰,
- [Christoph Scheiermann](#) ¹ &
- [Andrés Hidalgo](#) ²

Haematology has lost a giant: Paul Sylvain Frenette died in July, aged 56. His research led directly to the development of therapies that changed clinical practice. And he taught us — his former trainees — by example and shaped our careers.

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.

Additional access options:

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

Nature **597**, 31 (2021)

doi: <https://doi.org/10.1038/d41586-021-02357-1>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02357-1>

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

- CORRESPONDENCE
- 31 August 2021

Australian bush fires and fuel loads

- [Mark Adams](#) ⁰ &
- [Mathias Neumann](#) ¹

David Bowman and colleagues incorrectly cite our work to support their claim that politicians and the media misled the public by blaming Australia's 2019–20 wildfires on inappropriate land management ([D. Bowman *et al.* *Nature* 584, 188–191; 2020](#)).

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.

Additional access options:

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

Nature **597**, 31 (2021)

doi: <https://doi.org/10.1038/d41586-021-02358-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02358-0>

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

Work

- **Gender gap in US patents leads to few inventions that help women** [20 August 2021]
Career News • Inventor teams with women represent 16% of filed patents, yet are more likely than male teams to create women's health products.
- **Drowning in the literature? These smart software tools can help** [01 September 2021]
Technology Feature • Search engines that highlight key papers are keeping scientists up to date.
- **Using sound to explore events of the Universe** [30 August 2021]
Where I Work • Wanda Diaz Merced analyses gravitational waves from explosions of massive stars by converting them into 'chirps'.

- CAREER NEWS
- 20 August 2021

Gender gap in US patents leads to few inventions that help women

Inventor teams with women represent 16% of filed patents, yet are more likely than male teams to create women's health products.

- [Sara Reardon](#) ⁰



Women — especially women of colour — are under-represented on teams of inventors. Credit: Getty

When Rembrand Koning's wife was giving birth to their first child in 2017, the couple was dismayed by the lack of technology for new mothers who were recovering from pregnancy complications. Baby scales were hard to use one-handed, breast pumps were frustrating, and there were few resources to help Koning's wife with the post-partum pre-eclampsia that she experienced.

It seemed, Koning says now, as if designers of health-care technologies had little first-hand knowledge of childbirth and the difficulties that could arise.

Perhaps, reasoned Koning, an economist at Harvard University in Cambridge, Massachusetts, the relative scarcity of women's health products on the market was due to a scarcity of women inventing them. A study¹ that he published in June confirms this theory: few biotechnology patents are owned by women, and female inventors are significantly more likely than are male ones to patent health products for women.



The business of science

Patenting inventions has long been a mainly male endeavour: a 2016 report² by the Institute for Women's Policy Research, a US non-profit organization that aims to improve the lives of women by shaping public policy, found that a woman was cited as the lead inventor on just 7.7% of all patents filed

between 1977 and 2010 in the United States. Those female-owned patents tended to involve products such as jewellery and clothing.

Koning theorizes that those who have the opportunity to invent anything have a strong influence on what actually gets invented. When inventors are diverse, he adds, “those people may see problems that men may have dismissed or overlooked”.

To find out whether this trend extends to the health field, Koning and his colleagues looked at more than 400,000 biomedical patents filed between 1976 and 2010 with the US National Bureau of Economic Research, a non-profit organization that collates all data from the US Patent and Trademark Office. In 2010, they found that only 16.2% of patents were generated by research teams consisting mostly of women — up from 6.3% in 1976.

Next, using a machine-learning algorithm, the researchers scoured the patents for terms that indicate whether the invention related to women’s or men’s health, then determined the gender make-up of the patent-holders for each. Teams made up of all women, they found, were 35% more likely than all-male teams to invent technologies relating to women’s health. But teams made up of all women or all men were equally likely to patent technologies for men’s health. If women and men had produced an equal number of patents since 1976, the researchers estimated, there would be 6,500 more female-focused inventions today.

Unequal invention

Koning says that there are probably multiple reasons for these disparities. The unequal numbers of patents produced by men and women suggests that fewer women have the chance to invent. That’s not surprising, he says, given the barriers that women face advancing in science and engineering fields. Still, the discrepancy is particularly striking considering that [half of life-science PhDs in the United States are held by women](#).

Elina Berglund, a particle physicist and entrepreneur in New York City, is also not surprised by the findings. “It makes sense to me that one is more

likely to spend time innovating or solving a problem that they have either experienced or are close to,” she says.



Elina Berglund co-developed an app to track fertility cycles. Credit: Natural Cycles

After failing to find an effective non-hormone-based, non-invasive form of contraception, Berglund co-developed an app that used an algorithm to track fertility cycles. In 2018, Natural Cycles became the first direct-to-consumer contraceptive app to be approved by the US Food and Drug Administration. “I realized the unmet need for this option for contraception by experiencing it myself,” says Berglund, who is now chief executive of Natural Cycles. “I believe that the gender gap is indeed why women’s health has been an under-served and under-researched area for so long.”

Mathias Nielsen, a sociologist at the University of Copenhagen who researches social stratification in science, says that Koning’s study represents the first time that anyone has quantified an issue that he says researchers have long suspected: less attention is given to innovation in women’s health because the mostly male entrepreneurial field researches it

less. “It’s a very interesting and important contribution to the discussion of how diversity links to innovation,” he says.

Sociologist Laurel Smith-Doerr, who studies gender diversity in science, is not surprised by the gender gap in patenting, but says that the link between women’s patents and women’s health is an important advance. She thinks that further research is needed to examine how gender and ethnicity collectively influence invention trends. The same work needs to be repeated with nationality, she argues, and also with regards to whether the inventor is the first in their family to attend university.

Smith-Doerr, who is at the University of Massachusetts Amherst, adds that inventors are often drawn to work on problems that affect their own daily lives. That could partly explain why diseases that disproportionately affect those from minority ethnic groups or people who live in developing countries tend to receive less research funding and interest than do diseases that affect financially privileged white people. If research teams were more racially diverse, the focus of their interest might change, she says. “We’d be better off with inventions that represent all of the knowledge and all the human capacity we have,” she says.

Diversity, she adds, is also important for ensuring that medical devices and technologies that target all genders benefit all of them equally. “Historically, the male body has been taken as the baseline,” Nielsen says; certain technologies and practices such as personal protective equipment and drug-dosing regimens don’t necessarily work well for women.

Koning expects that the male-heavy invention trend is also visible in the concepts that make it to the marketplace. Venture capitalists are much less likely to fund women’s inventions: start-up companies led by women in the United States received [only 2.3% of total venture-capital funding in 2020](#), and women represent just 12% of venture-capital decision-makers. Koning thinks that this trend might extend to inventions that most benefit women. He cites stories about venture capitalists shying away from touching a breast pump, for instance, thinking it was “gross”. Men are also more likely to be on the scientific advisory boards of companies and start-up businesses³.

Koning says that his team is now looking into digital health products and apps, and whether they benefit certain communities or genders more than others. The team also plans to look at whether the gender imbalance in entrepreneurship affects the products that are brought to market. “Our hope is that in 20 years, some of this stuff starts disappearing,” he says.

Nature **597**, 139–140 (2021)

doi: <https://doi.org/10.1038/d41586-021-02298-9>

References

1. 1.

Koning, R., Samila, S. & Ferguson, J.-P. *Science* **372**, 1345–1348 (2021).

2. 2.

Milli, J., Gault, B., Williams-Baron, E., Xia, J. & Berlan, M. *The Gender Patenting Gap* (Institute for Women’s Policy Research, 2016).

3. 3.

Ding, W. W., Murray, F. & Stuart, T. E. *Acad. Manag. J.* **56**, 1443–1464 (2013).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02298-9>

- TECHNOLOGY FEATURE
- 01 September 2021

Drowning in the literature? These smart software tools can help

Search engines that highlight key papers are keeping scientists up to date.

- [David Matthews](#) ⁰



Illustration by The Project Twins

Every time Eddie Smolyansky had a few moments to himself, he tried to stay abreast of new publications in his field. But by 2016, the computer-

vision researcher, who is based in Tel Aviv, Israel, was receiving hundreds of automated literature recommendations per day. “At some point the bathroom breaks weren’t enough,” he says. The recommendations were “way too much, and impossible to keep up with”.

Smolyansky’s ‘feed fatigue’ will be familiar to many academics. Academic alert tools, originally designed to focus attention on relevant papers, have themselves become a hindrance, flooding the inboxes of scientists worldwide.

“I haven’t even been reading my automated PubMed searches lately because it really is overwhelming,” says Craig Kaplan, a biologist at the University of Pittsburgh in Pennsylvania. “I honestly cannot keep on top of the literature.”

But change is afoot. In 2019, Smolyansky co-founded Connected Papers, one of a new generation of visual literature-mapping and recommendation tools. Other services that promise to tame the information overload, integrating Twitter feeds and daily news as well as research, are also available.

Origin story

Instead of serving up a daily list of new articles by e-mail, Connected Papers uses a single, user-chosen ‘origin paper’ to build a map of related research, based partly on overlapping citations. The service recently surpassed one million users, Smolyansky says.

The maps are colour-coded by publication date, and users can toggle between ‘prior’, seminal, papers and later, ‘derivative’, works that build on them. The idea is that scientists can search for an origin paper that interests them, and see from the resulting map which recent papers have made a splash in their field, how they relate to other research, and how many citations they have accrued.



[tl;dr: this AI sums up research papers in a sentence](#)

“You do not have to sit on the hose of papers and look at every paper that comes out for fear of missing it,” says Smolyansky. The tool is also helpful when scientists want to dive into an entirely new field, he adds, providing an overview of the essential literature.

Another visual-mapping tool is Open Knowledge Maps, a service offered by a Vienna-based not-for-profit organization of the same name. It was founded in 2015 by Peter Kraker, a former scholarly-communication researcher at Graz University of Technology in Austria.

Open Knowledge Maps creates its maps based on keywords rather than a central article, and relies on text similarity and metadata to work out how papers are related. The tool arranges 100 papers in similar subfields into bubbles whose relative positions suggest similarity; a search for articles on ‘climate change’, for example, might yield a related bubble about ‘risk cognition’.

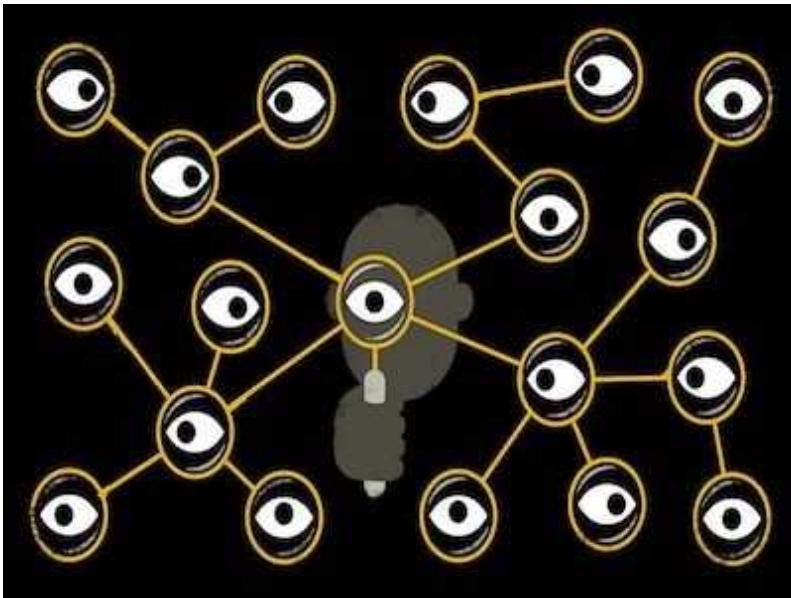
Maps of these bubbles can be built in about 20 seconds, and users can change them to include the 100 most recently published papers of relevance, or other resources. Open Knowledge Maps includes not only journal articles, but also content such as data sets and research software. Its users have created more than 400,000 maps so far, says Kraker.

Amie Fairs, who studies language at Aix-Marseille University in France, is a self-proclaimed Open Knowledge Maps enthusiast. “One particularly nice thing about Open Knowledge Maps is that you can search very broad topics, like ‘language production’, and it can group papers into themes you may not have considered,” Fairs says. For example, when she searched for ‘phonological brain regions’ — the areas of the brain that process sound and meaning — Open Knowledge Maps suggested a subfield of research about age-related differences in processing. “I hadn’t considered looking in the ageing literature for information about this before, but now I will,” she says.

Yet despite her enthusiasm for the service, Fairs still tends to find new papers through alerts from Google Scholar, the dominant tool in the field; it’s easier to go “down the rabbit hole”, she explains, following a chain of papers that cite each other.

Click to recommend

Google Scholar recommends papers depending on which articles users have authored and list in their profiles. The algorithm isn’t public, but the company says that the recommendations are based on “the topics of your articles, the places where you publish, the authors you work with and cite, the authors that work in the same area as you and the citation graph”. Users can manually set up extra e-mail alerts based on keyword searches or particular authors.



How AI technology can tame the scientific literature

Aaron Tay, a librarian at Singapore Management University who studies academic search tools, gets literature recommendations from both Twitter and Google Scholar, and finds that the latter often highlights the same articles as his human colleagues, albeit a few days later. Google Scholar “is almost always on target”, he says.

Besides published articles, Google Scholar might also pick up preprints as well as “low-quality theses and dissertations”, Tay says. Even so, “you get some gems you might not have seen”, he says. (Scopus, a competing literature database maintained by the Amsterdam-based publisher Elsevier, began incorporating preprints earlier this year, a spokesperson says. But it does not index theses and dissertations. “There will be titles that do not meet the Scopus standards but are covered by Google Scholar,” he says.)

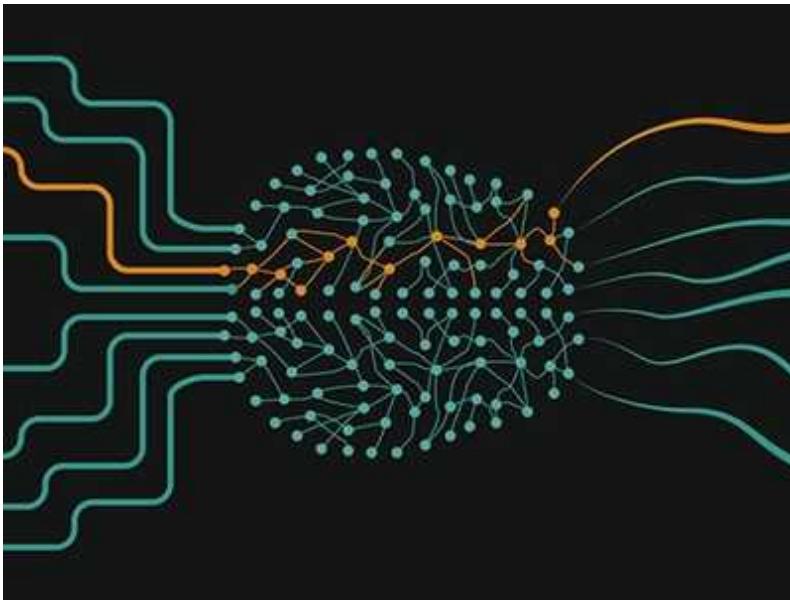
Google Scholar does not disclose the size of its database, but it is widely acknowledged to be the biggest corpus in existence, with close to 400 million articles by one estimate ([M. Gusenbauer *Scientometrics* 118, 177–214; 2019](#)). Open Knowledge Maps, meanwhile, is built on top of the open-source Bielefeld Academic Search Engine, which boasts more than 270 million documents, including preprints, and is curated to remove spam.

Connected Papers uses the publicly available corpus compiled by Semantic Scholar — a tool set up in 2015 by the Allen Institute for Artificial Intelligence in Seattle, Washington — amounting to around 200 million articles, including preprints. Smolyansky acknowledges this size discrepancy means that “very rarely” Google Scholar will find “some niche 1970s paper” that Semantic Scholar does not.

Semantic Scholar’s alert system, called an adaptive research feed, builds a list of recommended papers that users can train by liking or disliking the articles they see. To decide which papers are similar to those, it uses a machine-learning model trained on mutual citations, and on which articles Semantic Scholar users have viewed sequentially. It counts some 8 million monthly users.

No more FOMO

Feedly, launched in 2008, also uses upvotes and downvotes to learn which new academic research is most relevant to the user, and benefits from an AI assistant that can be trained on specific keywords or topics. But Feedly isn’t aimed specifically at researchers — it aims to be an all-encompassing dashboard to monitor news, RSS feeds (which provide a way of alerting users to new content on websites), the online forum Reddit, Twitter and podcasts. A free version is available, but extra features, such as the ability to follow more than 100 sources and hide adverts, cost US\$6 or more a month (unlike most of the other tools mentioned here, which are entirely free; another paid option is ResearchGate +Plus, which boosts users’ visibility and offers advanced statistics).



Artificial-intelligence tools aim to tame the coronavirus literature

ResearchRabbit, which fully launched in August 2021, describes itself as “Spotify for papers”. Users get started by saving relevant papers to a collection. With each added paper, ResearchRabbit updates its list of recommended articles, mirroring how the music-streaming platform makes recommendations based on the songs users add to their playlists. The company behind it, based in Seattle, Washington, hasn’t revealed exactly how it assesses relevance, although it says it focuses on precise recommendations rather than floods of alerts. “We only want to send the most relevant papers to our users,” says chief executive Michael Ma.

Amber Brown Ruiz, a special-education and disability-policy doctoral student at Virginia Commonwealth University in Richmond, finds ResearchRabbit alerts to be more personalized than Google Scholar, which sometimes feeds her papers that are superficially similar to her own work but turn out to be far outside her discipline.

Ruiz also uses Connected Papers to find new articles. She finds it to be less automated than Google Scholar, which sends fresh papers by e-mail, “but you can manually go in and figure out which articles are the newest”, she says.

What all these tools have in common is that they use some sort of artificial intelligence to craft their recommendations. But some scholars enjoy the human touch, valuing recommendations from colleagues and contacts on Twitter, for example. ResearchGate, the long-standing platform that brands itself as a kind of social network for scientists, says it offers the best of both worlds.



NatureTech hub

Founded in 2008, ResearchGate both e-mails recommendations of papers and serves them up through a rolling feed when users are logged in. (Users can also see a chronological newsfeed of papers posted by their ResearchGate contacts.) Although it doesn't make its algorithm public, it uses information about a user's publications and which publications they have viewed on the platform to understand their interests. It then calculates related articles on the basis of shared citations and extracted topics and keywords. ResearchGate currently includes some 149 million publication pages and has 20 million users.

“The secret sauce of ResearchGate is the combination of an active social network and a huge research graph,” says Joseph Debruin, ResearchGate’s director of product management, who is based in Los Angeles, California.

Five years after realizing he was drowning in new papers, Smolyansky is finally able to shake off that scientific ‘fear of missing out’. “You do not have to have that FOMO feeling,” he says.

Nature **597**, 141-142 (2021)

doi: <https://doi.org/10.1038/d41586-021-02346-4>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02346-4>

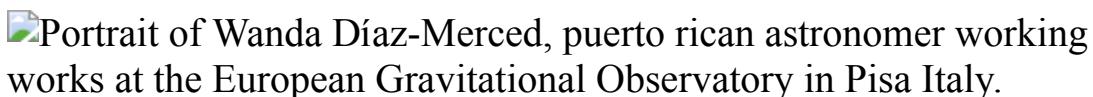
| [Section menu](#) | [Main menu](#) |

- WHERE I WORK
- 30 August 2021

Using sound to explore events of the Universe

Wanda Diaz Merced analyses gravitational waves from explosions of massive stars by converting them into ‘chirps’.

- [Davide Castelvecchi](#)



Wanda Diaz Merced is a computer scientist and astrophysicist with Virgo near Pisa, Italy. Credit: Enrico Sacchetti for *Nature*

I am standing in front of Virgo — a gravitational-wave observatory near Pisa, Italy, where I work as a computer scientist and astrophysicist — facing one of its 3-kilometre-long arms, in which a laser bounces between two mirrors in an internal pipe. The arms sense tiny changes in space-time (a model that fuses the three dimensions of space and the one dimension of time into a four-dimensional entity) when gravitational waves arrive from the distant Universe.

Some events in the Universe, such as massive stars exploding, produce gravitational waves — undulating space-time that propagates in all directions — and radiation that ordinary telescopes and X-ray telescopes can detect. Those emissions carry information about the origins of the events that produced them or about gravity itself.

I’m from Puerto Rico, and lost my sight in my twenties from degenerative diabetic retinopathy. So I use sound, as I’m doing here, to explore

measurements of gravitational waves. I earned a PhD in computer science at the University of Glasgow, UK, so that I could develop software tools to analyse astronomical data by converting it into sound. By ‘listening’ to the mathematics, I want to identify the possible X-ray counterparts to gravitational waves detected by Virgo.

Gravitational-wave data from colliding black holes is called a ‘chirp’: it sounds like a bird when it’s converted into sound.

Virgo, now closed for upgrades, will reopen after June 2022. Meanwhile, I’m working with colleagues on Research Infrastructures For Citizens in Europe. We’ll develop software that anyone, regardless of where they live or of any disabilities, can use to make multisensorial data and contribute to research.

To get a sense of Virgo’s size, I’ve walked to the end of one of its arms and back. I like to sing while walking to focus on my destination. But when the detector restarts, it is so sensitive that it might pick up my voice. So the people in the control room sometimes tell me I cannot walk and sing.

Nature **597**, 144 (2021)

doi: <https://doi.org/10.1038/d41586-021-02347-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02347-3>

Research

- **[Benzene rings broken for chemical synthesis](#)** [01 September 2021]

News & Views • Benzene rings are almost unbreakable in typical reaction conditions. Chemistry has now been developed that selectively breaks these rings open, highlighting their potential as building blocks for making open-chain molecules.

- **[cGAS-like receptors put a sting into the evolution of immune defence](#)** [09 August 2021]

News & Views • Aspects of how immune defence processes evolved remain mysterious. Studies of the fly *Drosophila melanogaster* reveal previously unknown details of a defence pathway with echoes of, but key differences from, a human pathway.

- **[Molecular memristors offer a path to ultra-efficient computing](#)** [01 September 2021]

News & Views • Future progress in computing calls for innovative ways to map the physical characteristics of materials to the logic functions needed by computing architectures. An electronic device called a molecular memristor provides a way forward.

- **[Insights into a receptor that lets insects sense scents](#)** [04 August 2021]

News & Views • In insects, odorant receptor proteins form membrane ion channels that open on binding to an odorant molecule. The structures of an inactive and an active channel lend insights into how insects detect and distinguish between odours.

- **[From the archive](#)** [31 August 2021]

News & Views • Nature's pages feature a look back at the life of Lawrence Bragg, and a call for London to erect a statue of Louis Pasteur.

- **[A brain signal that coordinates thought with metabolism](#)** [11 August 2021]

News & Views • In a brain structure called the hippocampus, sharp wave-ripples — oscillatory hallmarks of an ‘offline’ mode of cognitive processing — have been found to predict dips in glucose concentrations in the body.

- **[Measuring the density structure of an accretion hot spot](#)** [01 September 2021]

Article • An observed one-day difference between the peaks of emission of ultraviolet and optical light from the hot spot on GM Aurigae indicates that the hot spot has a radial density gradient.

- **Emergent hydrodynamics in a strongly interacting dipolar spin ensemble** [01 September 2021]

Article • A hybrid solid-state platform based on two strongly interacting dipolar species is used to study the emergence of the classical properties of a solid from its underlying microscopic quantum description.

- **Decision trees within a molecular memristor** [01 September 2021]

Article • Multiple redox transitions in a molecular memristor can be harnessed as ‘decision trees’ to undertake complex and reconfigurable logic operations in a single time step.

- **Scalable production of high-performing woven lithium-ion fibre batteries** [01 September 2021]

Article • Rechargeable lithium-ion batteries produced in the form of metre-long fibres can be woven into sturdy, washable textiles on an industrial loom and used to power other fabric-based electronic components.

- **Cleaving arene rings for acyclic alkenylnitrile synthesis** [19 July 2021]

Article • Common aromatic rings, such as anilines, arylboronic acids and aryl halides, can be opened up and converted to alkenyl nitriles through carbon–carbon bond cleavage using a copper catalyst.

- **Enantioselective synthesis of ammonium cations** [01 September 2021]

Article • Enantioselective supramolecular recognition allows for the asymmetric synthesis of nitrogen stereocentres, providing chiral ammonium cations in a dynamic crystallization process.

- **The contribution of insects to global forest deadwood decomposition** [01 September 2021]

Article • Multi-year field experiments across six continents suggest that insects have an important contribution to decomposition and carbon release from forest deadwood.

- **A metabolic function of the hippocampal sharp wave-ripple** [11 August 2021]

Article • Sharp wave-ripples from the hippocampus are shown to modulate peripheral glucose homeostasis in rats, offering insights into the mechanism that links sleep disruption and blood glucose regulation in type 2 diabetes.

- **Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche** [25 August 2021]

Article • Computational modelling and mouse genetics approaches show that multipotent progenitor cells that have the potential to populate the hepato-pancreato-biliary lineage tree persist in the pancreato-biliary organ rudiment.

- **APRIL limits atherosclerosis by binding to heparan sulfate proteoglycans** [25 August 2021]
Article • The heparan sulfate proteoglycan-binding cytokine APRIL has a protective role against atherosclerotic disease.
- **SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape** [14 July 2021]
Article • A survey of SARS-CoV-2 RBD antibodies identifies those with activity against diverse SARS-CoV-2 variants and SARS-related coronaviruses, highlighting epitopes and features to prioritize in antibody and vaccine development.
- **Broad sarbecovirus neutralization by a human monoclonal antibody** [19 July 2021]
Article • The human monoclonal antibody S2X259 cross-reacts with spike proteins from all clades of sarbecovirus, and provides prophylactic and therapeutic protection in vivo against parental SARS-CoV-2 and emerging variants of concern.
- **cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in Drosophila** [14 July 2021]
Article • The authors identify Drosophila cGLR1 as a double-stranded RNA sensor and 3'2'-cGAMP as a nucleotide second messenger and activator of Drosophila STING signalling.
- **Two cGAS-like receptors induce antiviral immunity in Drosophila** [14 July 2021]
Article • Two cGAS-like receptors, cGLR1 and cGLR2, identified in Drosophila melanogaster are shown to induce antiviral immunity in response to RNA or DNA virus infections through the production of 2'3'-cGAMP and 3'2'-cGAMP.
- **A clinically applicable integrative molecular classification of meningiomas** [25 August 2021]
Article • Multi-omics datasets are integrated to generate a unified and clinically informed molecular classification of meningiomas.
- **The structural basis of odorant recognition in insect olfactory receptors** [04 August 2021]
Article • Structural and functional analysis of an insect olfactory receptor shed light on how receptors can be activated by diverse odorants.
- **DAXX represents a new type of protein-folding enabler** [18 August 2021]
Article • A protein chaperone system is identified that consists of proteins with poly-Asp/Glu sequence, and may have an important role in diseases characterized by protein aggregation.

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

- NEWS AND VIEWS
- 01 September 2021

Benzene rings broken for chemical synthesis

Benzene rings are almost unbreakable in typical reaction conditions. Chemistry has now been developed that selectively breaks these rings open, highlighting their potential as building blocks for making open-chain molecules.

- [Mark R. Crimmin](#) ⁰

Synthetic chemists are always searching for new ways to make chemical bonds. Methods for forging carbon–carbon (C–C) bonds are especially important, because they enable organic molecules to be built from scratch, step by step and piece by piece. Once constructed, C–C bonds are often considered to be ‘untouchable’ parts of the molecule that generally resist further chemical transformation. [Writing in Nature](#), Qiu *et al.*¹ challenge this preconception by reporting a reaction that breaks the strong C–C bonds of organic molecules known as aromatic compounds.

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.

Additional access options:

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

Nature **597**, 33-34 (2021)

doi: <https://doi.org/10.1038/d41586-021-02322-y>

References

1. 1.
Qiu, X. *et al.* *Nature* **597**, 64–69 (2021).
2. 2.
Benitez, V. M., Grau, J. M., Yori, J. C., Pieck, C. L. & Vera, C. R. *Energy Fuels* **20**, 1791–1798 (2006).
3. 3.
Guengerich, F. P. & Yoshimoto, F. K. *Chem. Rev.* **118**, 6573–6655 (2018).
4. 4.
Hicks, J., Vasko, P., Goicoechea, J. M. & Aldridge, S. *J. Am. Chem. Soc.* **141**, 11000–11003 (2019).
5. 5.
Jakoobi, M., Halcovitch, N., Whitehead, G. F. S. & Sergeev, A. G. *Angew. Chem. Int. Edn* **56**, 3266–3269 (2017).

6. 6.

Hu, S., Shima, T. & Hou, Z. *Nature* **512**, 413–415 (2014).

7. 7.

Kong, R. Y. & Crimmin, M. R. *Angew. Chem. Int. Edn* **60**, 2619–2623 (2021).

8. 8.

Hall, J. H. *J. Am. Chem. Soc.* **87**, 1147–1148 (1965).

9. 9.

Kajimoto, T., Takahashi, H. & Tsuji, J. *J. Org. Chem.* **41**, 1389–1393 (1976).

10. 10.

Danishefsky, S. J. *J. Antibiot.* **72**, 319–322 (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02322-y>.

- NEWS AND VIEWS
- 09 August 2021

cGAS-like receptors put a sting into the evolution of immune defence

Aspects of how immune defence processes evolved remain mysterious. Studies of the fly *Drosophila melanogaster* reveal previously unknown details of a defence pathway with echoes of, but key differences from, a human pathway.

- [Cara West](#) ⁰ &
- [Neal Silverman](#) ¹

The detection of disease-causing agents by the rapidly responding branch of immune defences called the innate immune system relies on several classes of receptor. These detect molecular patterns associated with such agents, or the cellular damage that the agents cause. In particular, the surveillance of RNA and DNA is crucial to limit the replication of infectious organisms, especially intracellular microbes. Indeed, various sensors dedicated to the detection of nucleic acids have been characterized. Writing in *Nature*, [Slavik et al.](#)¹ and [Holleufer et al.](#)² describe surprising defence functions for cGAS-like receptors (cGLRs) of the fly *Drosophila melanogaster*, and reveal that this class of receptor is more diverse than was previously anticipated.

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.

Additional access options:

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

Nature **597**, 34-35 (2021)

doi: <https://doi.org/10.1038/d41586-021-02119-z>

References

1. 1.

Slavik, K. M. *et al.* *Nature* <https://doi.org/10.1038/s41586-021-03743-5> (2021).

2. 2.

Holleufer, A. *et al.* *Nature* <https://doi.org/10.1038/s41586-021-03800-z> (2021).

3. 3.

Crowl, J. T., Gray, E. E., Pestal, K., Volkman, H. E. & Stetson, D. B. *Annu. Rev. Immunol.* **35**, 313–336 (2017).

4. 4.

Ablasser, A. & Hur, S. *Nature Immunol.* **21**, 17–29 (2020).

5. 5.

Hopfner, K. P. & Hornung, V. *Nature Rev. Mol. Cell Biol.* **21**, 501–521 (2020).

6. 6.

Ahn, J. & Barber, G. N. *Exp. Mol. Med.* **51**, 1–10 (2019).

7. 7.

Goto, A. *et al. Immunity* **49**, 225–234 (2018).

8. 8.

Martin, M., Hiroyasu, A., Guzman, R. M., Roberts, S. A. & Goodman, A. G. *Cell Rep.* **23**, 3537–3550 (2018).

9. 9.

Liu, Y. & Cherry, S. *Autophagy* **15**, 174–175 (2019).

10. 10.

Cai, H. *et al. Sci. Signal* **13**, eabc4537 (2020).

11. 11.

Eaglesham, J. B., Pan, Y., Kupper, T. S. & Kranzusch, P. J. *Nature* **566**, 259–263 (2019).

- NEWS AND VIEWS
- 01 September 2021

Molecular memristors offer a path to ultra-efficient computing

Future progress in computing calls for innovative ways to map the physical characteristics of materials to the logic functions needed by computing architectures. An electronic device called a molecular memristor provides a way forward.

- [Matthew J. Marinella](#) ⁰ &
- [A. Alec Talin](#) ¹

Over the past 75 years, the performance, efficiency and information density of computers have each increased by a factor of about one trillion. Since the invention of the microchip, these improvements have largely resulted from the downscaling of electronic components known as complementary metal-oxide-semiconductor (CMOS) transistors. However, CMOS scaling has now lost steam and no longer provides marked performance upgrades, presenting a roadblock for computing progress. Overcoming this hurdle will require innovative methods to link the intrinsic physical properties of materials to the logic functions required by computing algorithms. In [a paper in *Nature*](#), Goswami *et al.*¹ demonstrate that the history-dependent electrical properties of a new molecular system can be harnessed to form both common and highly complex logic functions for a wide range of algorithms.

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.

Additional access options:

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

Nature **597**, 36-37 (2021)

doi: <https://doi.org/10.1038/d41586-021-02323-x>

References

1. 1.

Goswami, S. *et al.* *Nature* **597**, 51–56 (2021).

2. 2.

Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. *Nature* **453**, 80–83 (2008).

3. 3.

Balatti, S., Larentis, S., Gilmer, D. C. & Ielmini, D. *Adv. Mater.* **25**, 1474–1478 (2013).

4. 4.

Mickel, P. R., Lohn, A. J., James, C. D. & Marinella, M. J. *Adv. Mater.* **26**, 4486–4490 (2014).

5. 5.

Zhirnov, V. V. & Cavin, R. K. *Nature Mater.* **5**, 11–12 (2006).

6. 6.

van der Molen, S. J. & Liljeroth, P. *J. Phys. Condens. Matter* **22**, 133001 (2010).

7. 7.

Troisi, A., Orlandi, G. & Anthony, J. E. *Chem. Mater.* **17**, 5024–5031 (2005).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02323-x>

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

- NEWS AND VIEWS
- 04 August 2021

Insights into a receptor that lets insects sense scents

In insects, odorant receptor proteins form membrane ion channels that open on binding to an odorant molecule. The structures of an inactive and an active channel lend insights into how insects detect and distinguish between odours.

- [Emily R. Liman](#) [ORCID: http://orcid.org/0000-0003-4765-5496⁰](http://orcid.org/0000-0003-4765-5496)

As we apply insect repellents in an effort to thwart insect-borne diseases, we might stop to wonder how these substances work and why they are not more effective. DEET, the compound most commonly used in insect repellents, is thought to broadly activate insects' odorant receptor proteins, scrambling the olfactory code the insects use for seeking a host¹. But how DEET or natural odorant molecules bind to and affect the activity of insect odorant receptors has not been clear. [Writing in Nature](#), del Marmol *et al.*² report the structure of an insect odorant receptor in association with DEET or with the odorant eugenol, thereby providing key insights into how odorants bind to the receptor and how the structure of the activated receptor subsequently 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.

Additional access options:

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

Nature **597**, 37-39 (2021)

doi: <https://doi.org/10.1038/d41586-021-02076-7>

References

1. 1.

Pellegrino, M., Steinbach, N., Stensmyr, M. C., Hansson, B. S. & Vosshall, L. B. *Nature* **478**, 511–514 (2011).

2. 2.

del Mármol, J., Yedlin, M. A. & Ruta, V. *Nature* <https://doi.org/10.1038/s41586-021-03794-8> (2021).

3. 3.

Nara, K., Saraiva, L. R., Ye, X. & Buck, L. B. *J. Neurosci.* **31**, 9179–9191 (2011).

4. 4.

Araneda, R. C., Kini, A. D. & Firestein, S. *Nature Neurosci.* **3**, 1248–1255 (2000).

5. 5.
Hallem, E. A., Ho, M. G. & Carlson, J. R. *Cell* **117**, 965–979 (2004).
6. 6.
Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
7. 7.
Clyne, P. J. *et al.* *Neuron* **22**, 327–338 (1999).
8. 8.
Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. *Cell* **96**, 725–736 (1999).
9. 9.
Sato, K. *et al.* *Nature* **452**, 1002–1006 (2008).
10. 10.
Wicher, D. *et al.* *Nature* **452**, 1007–1011 (2008).
11. 11.
Fischer, E. *Ber. Deutsch. Chem. Ges. Berlin* **27**, 2985–2993 (1894).
12. 12.
Tripathi, A. & Bankaitis, V. A. *J. Mol. Med. Clin. Appl.* **2**, 10.16966/2575-0305.106 (2017).
13. 13.
Larsson, M. C. *et al.* *Neuron* **43**, 703–714 (2004).
14. 14.

Butterwick, J. A. *et al.* *Nature* **560**, 447–452 (2018).

15. 15.

Brand, P. *et al.* *eLife* **7**, e38340 (2018).

16. 16.

Catterall, W. A., Wisedchaisri, G. & Zheng, N. *Nature Chem. Biol.* **16**, 1314–1320 (2020).

17. 17.

Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. *J. Neurosci.* **25**, 1806–1815 (2005).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02076-7>

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

- NEWS AND VIEWS
- 31 August 2021

From the archive

Nature's pages feature a look back at the life of Lawrence Bragg, and a call for London to erect a statue of Louis Pasteur.

50 years ago

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.

Additional access options:

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

Nature **597**, 39 (2021)

doi: <https://doi.org/10.1038/d41586-021-02324-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02324-w>

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

- NEWS AND VIEWS
- 11 August 2021

A brain signal that coordinates thought with metabolism

In a brain structure called the hippocampus, sharp wave-ripples — oscillatory hallmarks of an ‘offline’ mode of cognitive processing — have been found to predict dips in glucose concentrations in the body.

- [Manfred Hallschmid](#) ⁰ &
- [Jan Born](#) ¹

To regulate adaptive behaviour, the brain relies on a continuous flow of cognitive and memory-related processes that require a constant energy supply. Weighing around 1,200 grams in women and 1,300 grams in men, on average, the brain consumes around 90 grams, or 340 kilocalories’ worth, of glucose per day, accounting for around half of the body’s glucose demand^{1,2}. The tight integration of metabolic and cognition-related signals might aid the matching of the brain’s energy supply to its energy needs, by optimizing foraging behaviour and efforts to limit energy expenditure. The synchronization of glucose supply with brain activity has so far been considered a function of a structure called the hypothalamus, at the base of the brain. [Writing in Nature](#), Tingley *et al.*³ provide evidence in rats for the role of another brain region, called the hippocampus, which is typically implicated in memory and navigation, in this equation (Fig. 1).

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.

Additional access options:

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

Nature **597**, 39-40 (2021)

doi: <https://doi.org/10.1038/d41586-021-02122-4>

References

1. 1.

Dienel, G. A. *Physiol. Rev.* **99**, 949–1045 (2019).

2. 2.

Wasserman, D. H. *Am. J. Physiol. Endocrinol. Metab.* **296**, E11–E21 (2009).

3. 3.

Tingley, D., McClain, K., Kaya, E., Carpenter, J. & Buzsáki, G. *Nature* <https://doi.org/10.1038/s41586-021-03811-w> (2021).

4. 4.

Buzsáki, G. *Hippocampus* **25**, 1073–1188 (2015).

5. 5.

Logothetis, N. K. *et al. Nature* **491**, 547–553 (2012).

6. 6.

Schwartz, M. W. *et al. Nature* **503**, 59–66 (2013).

7. 7.

Peters, A. *et al. Neurosci. Biobehav. Rev.* **28**, 143–180 (2004).

8. 8.

Chouhan, N. S., Griffith, L. C., Haynes, P. & Sehgal, A. *Nature* **589**, 582–585 (2021).

9. 9.

Biessels, G. J. & Despa, F. *Nature Rev. Endocrinol.* **14**, 591–604 (2018).

10. 10.

Tasali, E., Leproult, R., Ehrmann, D. A. & Van Cauter, E. *Proc. Natl Acad. Sci. USA* **105**, 1044–1049 (2008).

11. 11.

Higgs, S., Williamson, A. C., Rotshtein, P. & Humphreys, G. W. *Psychol. Sci.* **19**, 623–628 (2008).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-02122-4>

- Article
- Open Access
- [Published: 01 September 2021](#)

Measuring the density structure of an accretion hot spot

- [C. C. Espaillat](#) [ORCID: orcid.org/0000-0001-9227-5949¹](#),
- [C. E. Robinson](#) [ORCID: orcid.org/0000-0003-1639-510X²](#),
- [M. M. Romanova³](#),
- [T. Thanathibodee](#) [ORCID: orcid.org/0000-0003-4507-1710⁴](#),
- [J. Wendeborn](#) [ORCID: orcid.org/0000-0002-6808-4066¹](#),
- [N. Calvet](#) [ORCID: orcid.org/0000-0002-3950-5386⁴](#),
- [M. Reynolds⁴](#) &
- [J. Muzerolle](#) [ORCID: orcid.org/0000-0002-5943-1222⁵](#)

Nature volume 597, pages 41–44 (2021)

- 1219 Accesses
- 202 Altmetric
- [Metrics details](#)

Subjects

- [Astronomy and planetary science](#)
- [Astrophysical disks](#)

Abstract

Magnetospheric accretion models predict that matter from protoplanetary disks accretes onto stars via funnel flows, which follow stellar magnetic field lines and shock on the stellar surfaces^{1,2,3}, leaving hot spots with density gradients^{4,5,6}. Previous work has provided observational evidence of varying density in hot spots⁷, but these observations were not sensitive to the radial density distribution. Attempts have been made to measure this distribution using X-ray observations^{8,9,10}; however, X-ray

emission traces only a fraction of the hot spot^{11,12} and also coronal emission^{13,14}. Here we report periodic ultraviolet and optical light curves of the accreting star GM Aurigae, which have a time lag of about one day between their peaks. The periodicity arises because the source of the ultraviolet and optical emission moves into and out of view as it rotates along with the star. The time lag indicates a difference in the spatial distribution of ultraviolet and optical brightness over the stellar surface. Within the framework of a magnetospheric accretion model, this finding indicates the presence of a radial density gradient in a hot spot on the stellar surface, because regions of the hot spot with different densities have different temperatures and therefore emit radiation at different wavelengths.

[Download PDF](#)

Main

We conducted a coordinated multiepoch multiwavelength observing campaign of the accreting T Tauri star GM Aur, a roughly 2-Myr-old solar analogue that is surrounded by a disk with a large cavity with a radius of about 20 au^{15,16,17}. Gas is flowing inside the cavity and eventually reaches the star, as evidenced by moderate accretion rates¹⁸; however, most of the solid material in the cavity has grown to millimetre and larger sizes, as indicated by the lack of substantial infrared emission¹⁵ that would be present if a substantial amount of smaller dust grains remained. Because giant planets must be present for large cavities (tens of astronomical units in size^{19,20}) to open in protoplanetary disks, GM Aur is an ideal candidate to study a protoplanetary disk with the properties required to form planets.

Here we present results from our multiwavelength variability study, including Neil Gehrels Swift Observatory (Swift) X-ray and near-ultraviolet (NUV) fluxes, Hubble Space Telescope (HST) NUV spectra, Las Cumbres Observatory Global Telescope (LCOGT) $u'g'r'i'$ photometry, Transiting Exoplanet Survey Satellite (TESS) photometry and CHIRON H α spectra taken contemporaneously over roughly 35 days. We find no significant variation in the X-ray emission (Extended Data Fig. 1) and therefore rule it out as being responsible for any of the changes seen and do not discuss it further. Daily changes appear in the Swift NUV, LCOGT $u'g'r'i'$ and TESS light curves, with a period of about 6 days (Methods), consistent with the 6.1-day rotation period of the star²¹ and evidence of rotational modulation (Fig. 1). The NUV, u' , and g' fluxes (Fig. 1, red shaded boxes) peak about a day before (Methods) the TESS, r' and i' fluxes (Fig. 1, green shaded boxes). Specifically, the UV data peaked on 2019 December 1, 7, 13 (not as well defined owing to sparse data) and 19, whereas the optical data peaked on 2019 December 2, 8, 14 (not as well defined in the TESS light curve) and 20. There is also a dip in all the light curves on December 23–25 (Fig.

[1](#), grey shaded boxes). This is followed by the disappearance of the UV peaks in the light curves; the optical peaks remain, but appear to be lower than those before the dip in the light curves (Fig. [1](#)). We propose that this combination of features in the light curves is due to structure in the hot spot.

Fig. 1: Multiwavelength light curves of GM Aur.

 **figure1**

a–f, Swift NUV fluxes (**a**; Extended Data Table 1), LCOGT $u'g'$ flux calibrated magnitudes (**b, c**), TESS normalized magnitudes (**d**) and LCOGT $r'i'$ flux calibrated magnitudes (**e, f**). The TESS bandpass covers the bandpass of the r' and i' data. Uncertainties are smaller than the symbols. Red and green shaded boxes highlight

peaks in the UV and optical light curves, respectively, and are separated by the rotation period of the star. Dark grey shaded boxes highlight a contemporaneous dip in all the light curves.

The time lag between the peaks in the UV and optical light curves and the eventual disappearance of the UV peaks is the first observational evidence of a density gradient in a hot spot on a stellar surface. The UV emission traces the high-density region of the hot spot and the optical emission traces the low-density region⁷. As the star rotates, we see the different density regions of the hot spot as it goes in and out of view. When the high-density region of the hot spot disappears, the UV emission decreases substantially.

Accretion-shock modelling confirms that the high-density region of the hot spot dominates the UV emission and that the low-density region of the hot spot dominates the optical emission (Methods). These models consist of a vertical accretion column close to the surface of the star that is constrained by the stellar magnetic field and divided into three subregions: the pre-shock zone, the post-shock zone and the heated photosphere below the shock²². The top of the accretion column is met by the funnel flow and the base is the hot spot on the stellar surface. There is a peak in both the observed UV emission and the model emission from the high-density region of the hot spot about 1 day before the peak in both the observed optical emission and the model emission from the low-density region (Fig. 2). Periodicity in the light curves (Methods) consistent with the stellar rotation period²¹ points to rotational modulation. This supports our contention that we are seeing a hot spot with a density gradient as it rotates along with the star.

Fig. 2: Comparison of simulated hot-spot properties to observed light curves.

 figure2

Best-fitting accretion-shock model parameters (Extended Data Table 2) are derived by fitting HST spectra (Methods). **a, b**, TESS optical light curve (**a**) and LCOGT u' light curve (**b**) between 2019 December 6 and 10. **c, d**, When u' peaks, the total energy flux of the accretion column (**c**) and the contribution from the high-density region (**d**) are

highest. **e**, The size of the hot spot peaks along with the optical data. **f**, The low-density region dominates on all days, whereas the high-density region drops substantially after 2019 December 8.

The above is also consistent with magnetohydrodynamic simulations, which predict that there should be a density gradient in the accretion flow, column and hot spot^{5,8,23}. We quantitatively compare the accretion-shock modelling results to the size of the hot spot predicted by three-dimensional magnetohydrodynamic models⁵. The low-density region of the hot spot is predicted to cover 10%–20% of the stellar surface. At high-density, this coverage can drop to less than 1%. Our accretion-shock modelling shows that the surface coverage of the hot spot decreases at larger densities, consistent with the theoretical predictions. Specifically, the coverage of the low-density region is 10%–20% and the coverage of the high-density region is 0.1% (Fig. 2, Extended Data Table 2, Methods). The sizes of these accretion hot spots are also roughly consistent with those measured from Zeeman Doppler imaging of young stars²⁴.

We also qualitatively compare the observed light curves with three-dimensional magnetohydrodynamic simulations that have the approximate properties of GM Aur (Fig. 3, Methods). Our observations point to a hot spot with a density gradient where at times the smaller high-density region of the hot spot is no longer visible (presumably because it is behind the star) while the larger low-density region is still visible. This is consistent with three-dimensional magnetohydrodynamic simulations^{5,23,25}, which predict that there may be a density gradient in the hot spot and that at times parts of the hot spot may not be visible owing to stellar rotation (Fig. 4, top). The observations are also consistent with simulated light curves (Fig. 4, bottom). Like the observations, the simulated light curves are modulated mainly by the stellar rotation, and the emission from the high-density region of the spot peaks before the emission from the low-density region. Another piece of evidence of structure in the hot spot is the disappearance of the UV peaks in the light curves after December 25, while the optical peaks appear to remain (Fig. 1). The position, shape and structure of the hot spot varies with time, which particularly influences the high-density region of the hot spot, causing it to occasionally disappear (Fig. 4, Methods). Therefore, emission from the low-density region of the hot spot persists while emission from the high-density region of the hot spot may disappear, as observed here.

Fig. 3: Three-dimensional magnetohydrodynamic simulations of accretion.

 **figure3**

a, A three-dimensional view of matter at a density level of $5.3 \times 10^{-13} \text{ g cm}^{-3}$ (blue) flowing onto the star (yellow) along selected magnetic-field lines (red) at $3.13P_{\text{star}}$ (where P_{star} is the stellar rotation period). Distance is measured in stellar radii (R_{star}). Only the central region ($42R_{\text{star}}$) is shown. **b–d**, The corresponding hot spot (**b**), along with slices in the x – z (**c**) and x – y (**d**) planes; the colour scale denotes the density. Here μ is the magnetic moment, Ω is the rotation axis and ρ is the density.

Fig. 4: Simulations of the hot spot and light curves.

 **figure4**

a, The hot spot is crescent-shaped, is bent around the magnetic pole and has a density gradient. Here μ is the magnetic moment, Ω is the rotation axis, ρ is the density and t is time. **b**, The light curves generated by the less dense (solid) and densest (broken) regions of the hot spot are modulated by the stellar rotation period (P_{star}). The high-density region of the hot spot rotates out of view, leading to a time lag between the light curves at about $3P_{\text{star}}$ and $5P_{\text{star}}$; it may also occasionally disappear (at about $2P_{\text{star}}$ and $4P_{\text{star}}$).

The time delay between the peaks in the UV and optical data rules out the possibility that the hot spot is azimuthally symmetric. However, the simulations do not find that the hot spot is systematically asymmetric. Simulations show that the spots may

become asymmetric owing to complex processes at the disk–magnetosphere boundary (Fig. 4, Methods). The observed systematic behaviour may result from a more complex multipole magnetic field closer to the star, which redirects part of the funnel flow such that the spots become systematically asymmetric in the azimuthal direction^{26,27}. Previous three-dimensional simulations of stars with dipole and octupole fields show that if the octupole field dominates near the star it may redirect the flow and shape the hot spot²⁸.

There are dips in the Swift, LCOGT and/or TESS light curves on December 21–22 and December 23–24 that precede the disappearance of the UV peaks. The first dip on December 21–22 is narrow and seen only in the TESS data (the Swift and LCOGT observations did not overlap in time). We see the start of the second dip on December 23 in the TESS data and we trace the entire dip in the Swift and LCOGT data (Fig. 1, dark grey shaded boxes). Dust extinction may lead to dips in the light curve²⁹; however, GM Aur does not have substantial dust in the inner disk¹⁵. The observed dips are close to the expected minimum flux due to stellar rotation. There may have also been a decrease in accretion and hence hot-spot emission around this time, leading to a corresponding decrease at UV and optical wavelengths. This is consistent with the subsequent disappearance of the UV peaks and the lower optical peaks.

Methods

Observations and data reduction

Our campaign of GM Aur was undertaken mostly during 2019 December. Here we present data spanning the X-ray to the optical from Swift, HST, LCOGT, TESS and CHIRON. We provide details on the data reduction below.

Swift observations were taken with the X-ray Telescope³⁰ and the Ultraviolet and Optical Telescope³¹, using the UVM2 filter (2,221 Å), daily between 2019 November 27 and 2019 December 27, except on days when observations could not be scheduled. For the 27 observations, observation IDs, exposure times, start times, X-ray count rates, NUV fluxes and measurement uncertainties are listed in Extended Data Table 1. The observation IDs are as follows: 0034249045–0034249058, 0034249060–0034249063, 0034249065–0034249070 and 0034249072–0034249074. We used the High Energy Astrophysics Science Archive Research Center (HEASARC) HEASoft software (version 6.22.1) to measure count rates and NUV fluxes. X-ray emission from the stellar corona is known to be variable^{32,33}. X-ray count rates (0.5–10.0 keV) of GM Aur usually varied between 0.006 counts s⁻¹ and 0.029 counts s⁻¹ (Extended Data Fig. 1) and are comparable to those previously seen in GM Aur³⁴, pointing to an X-ray luminosity of around 3×10^{30} erg s⁻¹ (assuming a distance of around 160 pc³⁵)

throughout the campaign. The count rate was highest on 2019 November 27 (0.046 counts s⁻¹). None of the X-ray variability observed seems to correlate with any other data obtained during this campaign (Fig. 1, Extended Data Fig. 1) and so we do not discuss it further.

HST NUV–near-infrared (NIR) spectra (1,700–10,000 Å with a resolution, R , of 500–1,000) were taken with the Space Telescope Imaging Spectrograph (STIS) on 2019 December 6, 7, 8, 9 and 10 (Extended Data Fig. 2) in programme 16010. The spectra were obtained with the G230L, G430L and G750L gratings, with a 52'' × 2'' slit, within one orbit per visit. Data were reduced automatically by the HST pipeline; we correct fringing in the G750L data³⁶. Archival G230L and G430L spectra of the non-accreting star RECX 1 (programme 11616) were used as the template for the accretion-shock modelling analysis below. Typical uncertainties on the HST spectra are 3%–15%.

TESS data were taken on 2019 November 28 to 2019 December 23 (sector 19), with a cadence of 2 min (Fig. 1). The TESS bandpass covers 6,000–10,000 Å and is centred at 7,865 Å. Data were reduced using the Science Processing Operations Center pipeline. We used Lightkurve³⁷ to check the data quality and found little contamination from nearby sources. The gap in the data is due to data download during orbital perigee. Typical uncertainties in the TESS data are 0.25%.

$u'g'r'i'$ data were taken roughly 5–10 times on clear nights at the LCOGT³⁸, with the Sinistro Imagers on the 1-m telescopes, between 2019 November 26 and 2020 January 1. The $u'g'r'i'$ filters have central wavelengths of 3,540 Å, 4,770 Å, 6,215 Å and 7,545 Å, and wavelength widths of 570 Å, 1,500 Å, 1,390 Å and 1,290 Å. Data in Fig. 1 were reduced using the Aperture Photometry Tool and standard aperture photometry techniques. Uncertainties in all bands are less than 0.005 mag. The flux of GM Aur was calibrated using the fluxes of background objects, which in turn were flux-calibrated using UBVRI data of background objects (converted to $u'g'r'i'$ ³⁹) and the GD 64 standard field⁴⁰. These calibration data were taken contemporaneously at the 4.3-m Lowell Discovery Telescope (LDT), using the Large Monolithic Imager, on 2019 December 2, 7, 10, 13, 18 and 21. The LDT data are used here only to flux-calibrate the LCOGT data.

Medium-resolution ($R \approx 25,000$) optical (4,082–8,906 Å) spectra were obtained with CHIRON⁴¹, on a 1.5-m telescope that is part of the Small and Moderate Aperture Research Telescope System (SMARTS) at Cerro Tololo Inter-American Observatory, between 2019 November 28 and December 17 (Extended Data Fig. 3). The standard CHIRON pipeline is not optimized to extract the H α profile of young stars; we reduce the spectra here⁴².

Timing analysis

Although the periodicity and time lag in the light curves are evident (Fig. 1), here we measure the period of the TESS and LCOGT $u'g'r'i'$ light curves as well as the time lag between the peak in the u' light curve and those in the $g'r'i'$ and TESS light curves.

The Swift NUV data are much sparser, and we do not analyse them in as much detail.

To measure the period in the light curves, we use the Astropy Lomb–Scargle periodogram function^{43,44,45}. We obtain a 5.8-day period for the TESS light curve; for the $u'g'r'i'$ light curves, we measure periods of 6.3, 6.3, 6.3 and 6.1 days. The Swift NUV flux is greater than 50% higher than surrounding days on November 30–December 1, December 7–8 and December 19 (Fig. 1), which is consistent with a roughly 6-day period. Given that the independently measured rotation period of the star is 6.1 days²¹, we attribute the roughly 6-day period measured in the light curves to stellar rotation.

For the time-lag analysis, we use the Python package Stingray⁴⁶. When compared to the u' light curve, we measure time lags of around 0, 0.75, 1.1 and 1 days for the $g'r'i'$ and TESS light curves. The Swift NUV flux is highest on the same days as the peaks seen in the u' light curve (Fig. 1). We conclude that the time lag between the UV and optical data is about 1 day.

Accretion-shock modelling

To further explore the roughly 1-day time lag in the peak between the UV and optical data seen in Fig. 1, we focus on 2019 December 6–10 (Fig. 2a,b) because HST data were taken daily during this time. HST NUV spectra are the best measure of accretion because the accretion column that channels material onto the surface of the star emits substantial energy at NUV wavelengths¹.

The HST data were fitted using accretion-shock models²². These models provide information on the physical properties of the accretion column and the associated hot spot, which is the footprint of the accretion column on the stellar surface. The accretion column is characterized by an energy flux ($\rho v_s^3/2$), which measures the density of material in the accretion column (ρ), assuming that the magnetospheric radius (R_{mag}) and infall velocity (v_s ; here 456 km s^{-1} , which depends on the stellar radius R_{star} , stellar mass M_{star} and R_{mag}) are constant. Each column has a filling factor f , which gives the fraction of the stellar surface covered by the column (that is, the hot spot).

GM Aur is a well studied source in the NUV, with eight previously modelled epochs¹⁸. In Fig. 2c–f, we plot the accretion-column model parameters (Extended Data Table 2) obtained from fitting the HST data (Extended Data Fig. 2). The fitting was done with three accretion columns^{7,18}, with energy fluxes of $1 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$, $1 \times 10^{11} \text{ erg s}^{-1} \text{ cm}^{-2}$ and $1 \times 10^{12} \text{ erg s}^{-1} \text{ cm}^{-2}$, and adopting published stellar parameters⁴⁷. Here we refer to the $1 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$ and $1 \times 10^{12} \text{ erg s}^{-1} \text{ cm}^{-2}$ components as the ‘low-density’ and ‘high-density’ regions. The total energy flux (F_{total} ; Fig. 2c) is the sum of the energy fluxes of all the regions, weighted by their respective f . The total hot-spot coverage on the stellar surface (Fig. 2e) is the sum of the f values for all the regions.

The total energy flux peaks along with the UV data (Fig. 2c). When the energy flux peaks on December 7, it is dominated by emission from a high-density region, which then drops substantially (Fig. 2d). This suggests that on December 7 we see the high-density region of the hot spot and that in the following three days most of this high-density region is no longer visible. Meanwhile, the hot-spot coverage of the star peaked about 1 day later, on December 8, along with the optical data (Fig. 2e), and the hot spot is dominated by emission from a low-density region (Fig. 2f). Because the low-density region of the accretion column emits its energy at longer wavelengths⁷, it follows that the optical emission peaks when the hot spot is at its largest. The observed behaviour in the light curves may be interpreted by combining two effects: the stellar rotation and the different physical locations for the high-density and low-density regions. It then follows that the high-density region of the hot spot (whose energy appears mainly in the UV) will lead to a peak in the UV data when the high-density parts of the hot spot are visible to the observer. As the star rotates, we may be seeing a denser part of the hot spot first, which then rotates out of view.

Accretion-flow modelling

To facilitate comparison to the three-dimensional (3D) magnetohydrodynamic (MHD) simulations, we need to estimate the magnetospheric radius of GM Aur.

Magnetospheric accretion-flow modelling^{48,49,50} of the H α line provides properties of the accretion flow. The model assumes that the magnetic, stellar-rotation and disk-rotation axes are aligned. The material flows onto the star along an axisymmetric accretion flow that arises from the co-rotating gas disk. The geometry of the flow is described by a dipolar magnetic field and characterized by an inner radius (R_i , which corresponds to R_{mag}) and the width of the flow (W_r) at the disk plane. The model assumes a steady-flow prescription for a given accretion rate to determine the density at a given point. The temperature at each point is determined parametrically, scaled to the density assuming a constant heating rate in the flow; the maximum temperature in the flow (T_{max}) describes each model. To calculate the emission line profile, the model

assumes the extended Sobolev approximation and calculates the mean intensity and the level population of a 16-level hydrogen atom, and uses the ray-by-ray method for a given viewing inclination (i).

We created a large grid of models varying the accretion rate, R_{mag} , W_r , T_{max} and i , using the ranges of parameters appropriate for accreting T Tauri stars⁴⁹. With these combinations, we calculated around 72,000 model profiles. We convolved the model profiles with a Gaussian instrumental profile of CHIRON’s resolution and fitted each observed profile inside $\pm 400 \text{ km s}^{-1}$ from the line centre. The best fits are determined by calculating the χ^2 for each combination of the model and observed profile.

For each observed profile, we selected 100 best fits and calculated the means of the accretion rate, R_{mag} , W_r , T_{max} and i (Extended Data Table 3). The accretion rates listed in Extended Data Table 3 are higher on the dates for which there are peaks in the UV emission. The derived i are roughly consistent with the measured inclination of the disk, which is 53° ¹⁷. W_r varies between $0.2R_{\text{star}}$ and $0.5R_{\text{star}}$; T_{max} varies between 8,270 K and 9,120 K. We use the derived R_{mag} of GM Aur to compare it to simulations in the next section.

In Extended Data Fig. 3, we show the best-fitting model to the H α profiles. The fit to the line wings is very good. The regions with strong absorption features on the blue side of the line were excluded from the fit; this blueshifted absorption is probably from winds, which are not included in the model. There is no periodic pattern in the H α line (Extended Data Fig. 3). However, multicomponent high-velocity blueshifted absorption, starting on December 7, occurs along with peaks in the UV emission (Fig. 1) and accretion rate (Extended Data Tables 2, 3). There is a possible second component at higher, blueshifted velocity on December 6, which might imply a deceleration after the launching of a higher-density outflow event that may presage the accretion event on December 7. Models⁵¹ show that there is a disk-wind component to the H α profile, which manifests as blueshifted absorption; this component is dominated by the innermost disk, within tens of stellar radii.

3D MHD simulations of accretion

Here we show global 3D MHD simulations of a rotating magnetized star accreting from a disk^{5,23,25}. In brief, the models assume that the star has a dipole magnetic field with a misalignment angle (θ) between the rotation axis (Ω) and magnetic axis (μ). The rotation axes of the star and disk are aligned. Here we use simulations with the same set up as previous work^{23,25}, with parameters chosen to approximate the properties of GM Aur.

We assume that the magnetic field of GM Aur can be approximated with a dipole field. The hot spot of GM Aur is at a high latitude of about 77° , measured from radial velocity variations of He i (5,876 Å)⁵²; previous work also found that the hot spot is at high latitudes for dipole–octupole configurations¹. Also, GM Aur has a long rotation period²¹, and young stars with simpler, more dipole magnetic fields are slower rotators⁵³. Simulations²³ show that the properties of the hot spot depend on θ , the corotation radius (R_{co}) and R_{mag} . For GM Aur, $\theta = 13^\circ$ ⁵². This is consistent with the range of inclinations of the accretion flow inferred from modelling the Hα profile (all the best-fitting inclinations in Extended Data Table 3 are within about 10° of the system inclination of 53°). Using a stellar rotation period of 6.1 days²¹, $R_{\text{star}} = 2R_{\text{sun}}$ and $M_{\text{star}} = 1.36M_{\text{sun}}$ ⁴⁷ (where R_{sun} and M_{sun} are the radius and mass of the Sun), $R_{\text{co}} = 7.8R_{\text{star}}$ (0.07 au). We measure a mean R_{mag} of about $3.8R_{\text{star}}$, with a range of $3.4R_{\text{star}}$ to $4.6R_{\text{star}}$ (Extended Data Table 3). The simulations use $\theta = 20^\circ$, $R_{\text{mag}} \approx 4.5R_{\text{star}}$ and $R_{\text{co}} = 5.7R_{\text{star}}$, which are consistent with the parameters of GM Aur.

The 3D MHD simulations have an approximate accretion rate of about $1.1 \times 10^{-8}M_{\text{sun}} \text{ yr}^{-1}$, which agrees with measurements from the accretion-shock and accretion-flow modelling (Extended Data Tables 2, 3). The energy distribution in the spots in the 3D MHD simulations varies between about $3.9 \times 10^9 \text{ erg s}^{-1} \text{ cm}^{-2}$ and $1.2 \times 10^{11} \text{ erg s}^{-1} \text{ cm}^{-2}$; this is consistent with the accretion-shock models, which use energy fluxes of $1 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$ to $1 \times 10^{12} \text{ erg s}^{-1} \text{ cm}^{-2}$. The densities of the 3D MHD simulated spots in Fig. 4 range between about $5.5 \times 10^{-13} \text{ g cm}^{-3}$ and $5.97 \times 10^{-12} \text{ g cm}^{-3}$, which overlaps with the accretion-shock models, where the densities are $2.1 \times 10^{-13} \text{ g cm}^{-3}$ to $2.1 \times 10^{-11} \text{ g cm}^{-3}$. The simulated light curves peak at a luminosity of about $1 \times 10^{31} \text{ erg s}^{-1}$ (Fig. 4), which is roughly consistent with the accretion luminosities of about $0.1L_{\text{sun}}$ to $0.2L_{\text{sun}}$ ($3.8 \times 10^{32} \text{ erg s}^{-1}$ to $7.7 \times 10^{32} \text{ erg s}^{-1}$; L_{sun} is the luminosity of the Sun) measured from the HST spectra. In Fig. 4 we show the light curve only once accretion onto the star begins in the simulation. To generate the light curve for the densest region of the hot spot, we use $\rho > 6.5 \times 10^{-12} \text{ g cm}^{-3}$.

A ratio of $R_{\text{co}}/R_{\text{mag}} = 1.5$ sets the boundary between stable and unstable accretion²⁵. In the stable regime, matter accretes onto the star in ordered funnel streams, and symmetric, crescent-shaped hot spots are expected⁵⁴. In the unstable regime, matter accretes in chaotic hot spots²⁵. The simulations have $R_{\text{co}}/R_{\text{mag}} = 1.3\text{--}1.4$. The measured R_{mag} of GM Aur has an uncertainty of about $1R_{\text{star}}$, so $R_{\text{co}}/R_{\text{mag}}$ for GM Aur could reach 1.4. Therefore, the simulations are consistent with the properties of GM Aur, and both are near the boundary of the stable–unstable regime. Here, processes at

the disk and magnetosphere boundary may cause the behaviour seen in the observed light curves. The inner disk rotates more rapidly than the magnetosphere, leading to non-stationary behaviour of matter in the inner disk and corresponding non-stationary behaviour of the hot spot. The hot spots are predominantly crescent-shaped, but may become asymmetric (Fig. 4). Also resulting from the difference in the disk and magnetosphere rotation, the non-stationarity leads to variation of the density distribution in the hot spot that especially influences the densest parts of the spot, which may occasionally disappear. This may explain why the UV emission (associated with the high-density region of the hot spot) disappears in Fig. 1.

Data availability

The raw data for the Swift, HST, TESS (<https://archive.stsci.edu>) and LCOGT (<https://lco.global/>) data are publicly available from their respective archives. The LDT and CHIRON raw data are available on request from C.C.E. HST and TESS provide reduced data in their archives. Reduced Swift, LCOGT, CHIRON and LDT data are available on request from C.C.E. Data from the accretion-shock and accretion-flow modelling are available on request from C.C.E. Data relevant to the 3D MHD numerical simulations are available on request from M.M.R. (romanova@astro.cornell.edu).

Code availability

Access to the accretion-shock, accretion-flow and MHD model codes are available on request from N.C. (ncalvet@umich.edu), J.M. (muzerol@stsci.edu) and M.M.R., respectively.

References

1. 1.

Hartmann, L., Herczeg, G. & Calvet, N. Accretion onto pre-main-sequence stars. *Annu. Rev. Astron. Astrophys.* **54**, 135–180 (2016).

2. 2.

Bouvier, J. et al. Probing the magnetospheric accretion region of the young pre-transitional disk system DoAr 44 using VLTI/GRAVITY. *Astron. Astrophys.* **636**, A108 (2020).

3. 3.

GRAVITY Collaboration. A measure of the size of the magnetospheric accretion region in TW Hydreae. *Nature* **584**, 547–550 (2020).

4. 4.

Romanova, M. M. & Owocki, S. P. Accretion, outflows, and winds of magnetized stars. *Space Sci. Rev.* **191**, 339–389 (2015).

5. 5.

Romanova, M. M., Ustyugova, G. V., Koldoba, A. V. & Lovelace, R. V. E. Three-dimensional simulations of disk accretion to an inclined dipole. II. Hot spots and variability. *Astrophys. J.* **610**, 920–932 (2004).

6. 6.

Lamzin, S. A. The structure of shock waves in the case of accretion onto low-mass young stars. *Astron. Rep.* **42**, 322–335 (1998).

7. 7.

Ingleby, L. et al. Accretion rates for T Tauri stars using nearly simultaneous ultraviolet and optical spectra. *Astrophys. J.* **767**, 112 (2013).

8. 8.

Bonito, R. et al. Magnetohydrodynamic modeling of the accretion shocks in classical T Tauri stars: the role of local absorption in the X-ray emission. *Astrophys. J.* **795**, L34 (2014).

9. 9.

Argiroffi, C. et al. Variable X-ray emission from the accretion shock in the classical T Tauri star V2129 Ophiuchi. *Astron. Astrophys.* **530**, A1 (2011).

10. 10.

Schneider, P. C. et al. Multiepoch, multiwavelength study of accretion onto T Tauri. X-ray versus optical and UV accretion tracers. *Astron. Astrophys.* **618**, A55 (2018).

11. 11.

Sacco, G. G. et al. On the Observability of T Tauri accretion shocks in the X-ray band. *Astron. Astrophys.* **522**, A55 (2010).

12. 12.

Orlando, S. et al. X-ray emitting MHD accretion shocks in classical T Tauri stars. Case for moderate to high plasma-beta values. *Astron. Astrophys.* **510**, A71 (2010).

13. 13.

Brickhouse, N. S. et al. A deep Chandra X-ray spectrum of the accreting young star TW Hydrae. *Astrophys. J.* **710**, 1835–1847 (2010).

14. 14.

Dupree, A. K. et al. TW Hya: spectral variability, X-rays, and accretion diagnostics. *Astrophys. J.* **750**, 73 (2012).

15. 15.

Calvet, N. et al. Disks in transition in the Taurus population: Spitzer IRS spectra of GM Aurigae and DM Tauri. *Astrophys. J.* **630**, L185–L188 (2005).

16. 16.

Andrews, S. M. et al. Resolved images of large cavities in protoplanetary transition disks. *Astrophys. J.* **732**, 42 (2011).

17. 17.

Macías, E. et al. Multiple rings in the transitional disk of GM Aurigae revealed by VLA and ALMA. *Astrophys. J.* **865**, 37 (2018).

18. 18.

Robinson, C. E. & Espaillat, C. C. Multiepoch ultraviolet HST observations of accreting low-mass stars. *Astrophys. J.* **874**, 129 (2019).

19. 19.

Zhu, Z., Nelson, R. P., Hartmann, L., Espaillat, C. & Calvet, N. Transitional and pre-transitional disks: gap opening by multiple planets? *Astrophys. J.* **729**, 47 (2011).

20. 20.

Espaillat, C. et al. in *Protostars and Planets VI* (eds Beuther, H. et al.) 497–520 (Univ. Arizona Press, 2014).

21. 21.

Percy, J. R., Grynko, S., Seneviratne, R. & Herbst, W. Self-correlation analysis of the photometric variability of T Tauri stars. II. A survey. *Publ. Astron. Soc. Pacif.* **122**, 753–765 (2010).

22. 22.

Calvet, N. & Gullbring, E. The structure and emission of the accretion shock in T Tauri stars. *Astrophys. J.* **509**, 802–818 (1998).

23. 23.

Kulkarni, A. K. & Romanova, M. M. Accretion to magnetized stars through the Rayleigh–Taylor instability: global 3D simulations. *Mon. Not. R. Astron. Soc.* **386**, 673–687 (2008).

24. 24.

Donati, J.-F. et al. Non-stationary dynamo and magnetospheric accretion processes of the classical T Tauri star V2129 Oph. *Mon. Not. R. Astron. Soc.* **412**, 2454–2468 (2011).

25. 25.

Blinova, A. A., Romanova, M. M. & Lovelace, R. V. E. Boundary between stable and unstable regimes of accretion. Ordered and chaotic unstable regimes. *Mon. Not. R. Astron. Soc.* **459**, 2354–2369 (2016).

26. 26.

Donati, J.-F. et al. Magnetic fields and accretion flows on the classical T Tauri star V2129 Oph. *Mon. Not. R. Astron. Soc.* **380**, 1297–1312 (2007).

27. 27.

Gregory, S. G., Matt, S. P., Donati, J.-F. & Jardine, M. The non-dipolar magnetic fields of accreting T Tauri stars. *Mon. Not. R. Astron. Soc.* **389**, 1839–1850 (2008).

28. 28.
Romanova, M. M., Long, M., Lamb, F. K., Kulkarni, A. K. & Donati, J.-F. Global 3D simulations of disc accretion on to the classical T Tauri star V2129 Oph. *Mon. Not. R. Astron. Soc.* **411**, 915–928 (2011).
29. 29.
Stauffer, J. et al. CSI 2264: characterizing young stars in NGC 2264 with short-duration periodic flux dips in their light curves. *Astron. J.* **149**, 130 (2015).
30. 30.
Burrows, D. N. et al. The Swift X-ray telescope. *Space Sci. Rev.* **120**, 165–195 (2005).
31. 31.
Roming, P. W. A. et al. The Swift ultra-violet/optical telescope. *Space Sci. Rev.* **120**, 95–142 (2005).
32. 32.
Preibisch, T. et al. The origin of T Tauri X-ray emission: new insights from the Chandra Orion ultradeep project. *Astrophys. J. Suppl. Ser.* **160**, 401–422 (2005).
33. 33.
Guarcello, M. G. et al. CSI 2264: simultaneous optical and X-ray variability in pre-main sequence stars. I. Time resolved X-ray spectral analysis during optical dips and accretion bursts in stars with disks. *Astron. Astrophys.* **602**, A10 (2017).
34. 34.
Espaillat, C. C., Robinson, C., Grant, S. & Reynolds, M. Using multiwavelength variability to explore the connection among X-ray emission, the far-ultraviolet H₂ bump, and accretion in T Tauri stars. *Astrophys. J.* **876**, 121 (2019).
35. 35.
Gaia Collaboration. Gaia data release 2. Summary of the contents and survey properties. *Astron. Astrophys.* **616**, A1 (2018).
36. 36.

Goudfrooij, P., & Christensen, J. A. *STIS Near-IR Fringing. III. A Tutorial on the Use of the IRAF Tasks*. Report No. 98-29 (STIS Instrument Science, 1998).

37. 37.

Lightkurve Collaboration. Lightkurve: Kepler and TESS time series analysis in Python. *Astrophysics Source Code Library* 1812.013 (2018).

38. 38.

Brown, T. M. et al. Las Cumbres Observatory global telescope network. *Publ. Astron. Soc. Pacif.* **125**, 1031 (2013).

39. 39.

Jester, S. et al. The Sloan Digital Sky Survey view of the Palomar-Green bright quasar survey. *Astron. J.* **130**, 873–895 (2005).

40. 40.

Landolt, A. U. UVRI photometric standard stars around the sky at +50 deg declination. *Astron. J.* **146**, 131 (2013).

41. 41.

Schwab, C., Spronck, J. F. P., Tokovinin, A. & Fischer, D. A. Design of the CHIRON high-resolution spectrometer at CTIO. *Proc. SPIE* **7735**, 77354G (2010).

42. 42.

Walter, F. *An Alternative Scheme for Chiron Fiber and Slicer Mode Data Reductions*
http://www.astro.sunysb.edu/fwalter/SMARTS/CHIRON/ch_reduce.pdf (2008).

43. 43.

Astropy Collaboration. The Astropy project: building an open-science project and status of the v2.0 core package. *Astron. J.* **156**, 123 (2018).

44. 44.

Lomb, N. R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39**, 447–462 (1976).

45. 45.

Scargle, J. D. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).

46. 46.

Huppenkothen, D. et al. Stingray: a modern Python library for spectral timing. *Astrophys. J.* **881**, 39 (2019).

47. 47.

Manara, C. F. et al. Gas content of transitional disks: a VLT/X-Shooter study of accretion and winds. *Astron. Astrophys.* **568**, A18 (2014).

48. 48.

Hartmann, L., Hewett, R. & Calvet, N. Magnetospheric accretion models for T Tauri stars. I. Balmer line profiles without rotation. *Astrophys. J.* **426**, 669–687 (1994).

49. 49.

Muzerolle, J., Calvet, N. & Hartmann, L. Magnetospheric accretion models for the hydrogen emission lines of T Tauri stars. *Astrophys. J.* **492**, 743–753 (1998).

50. 50.

Muzerolle, J., Calvet, N. & Hartmann, L. Emission-line diagnostics of T Tauri magnetospheric accretion. II. Improved model tests and insights into accretion physics. *Astrophys. J.* **550**, 944–961 (2001).

51. 51.

Lima, G. H. R. A., Alencar, S. H. P., Calvet, N., Hartmann, L. & Muzerolle, J. Modeling the H α line emission around classical T Tauri stars using magnetospheric accretion and disk wind models. *Astron. Astrophys.* **522**, A104 (2010).

52. 52.

McGinnis, P., Bouvier, J. & Gallet, F. The magnetic obliquity of accreting T Tauri stars. *Mon. Not. R. Astron. Soc.* **497**, 2142–2162 (2020).

53. 53.

Johnstone, C. P. et al. Tauri stars: magnetic fields, coronae, and star-disc interactions. *Mon. Not. R. Astron. Soc.* **437**, 3202–3220 (2014).

54. 54.

Kulkarni, A. K. & Romanova, M. M. Analytical hotspot shapes and magnetospheric radius from 3D simulations of magnetospheric accretion. *Mon. Not. R. Astron. Soc.* **433**, 3048–3061 (2013).

Acknowledgements

C.C.E. acknowledges support from HST grant GO-16010, NASA grant 80NSSC19K1712 and NSF career grant AST-1455042. M.M.R. acknowledges the NSF grant AST-20009820. N.C. acknowledges support from NASA grant NNX17AE57G.

Author information

Affiliations

1. Institute for Astrophysical Research, Department of Astronomy, Boston University, Boston, MA, USA

C. C. Espaillat & J. Wendeborn

2. Department of Physics and Astronomy, Amherst College, Amherst, MA, USA

C. E. Robinson

3. Department of Astronomy, Cornell University, Ithaca, NY, USA

M. M. Romanova

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

T. Thanathibodee, N. Calvet & M. Reynolds

5. Space Telescope Science Institute, Baltimore, MD, USA

J. Muzerolle

Contributions

C.C.E. planned the observing campaign, analysed and interpreted the data, and wrote the manuscript. C.E.R. reduced the HST, TESS and CHIRON data, provided related text and did the HST data modelling. M.M.R. did the 3D MHD simulations, provided related text and aided in the interpretation of the data. T.T. did the CHIRON data modelling and provided related text. J.W. reduced the LCOGT and LDT data and did the timing analysis. N.C. aided in the interpretation of the data. M.R. reduced the Swift data. J.M. aided in the interpretation of the CHIRON data. All authors participated in the discussion of results and revision of the manuscript.

Corresponding author

Correspondence to [C. C. Espaillat](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Hans Moritz Guenther 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 X-ray count rates of GM Aur.](#)

Data were observed by Swift between 2019 November 26 and 2020 January 1, and are listed in Extended Data Table 1. To facilitate comparison, we use the same time range and colour bars as in Fig. 1 for the Swift NUV data. Error bars correspond to uncertainties, which are listed in Extended Data Table 1.

[Extended Data Fig. 2 Accretion shock modelling.](#)

We show the best-fitting models (red solid line) to HST NUV–NIR spectra (black solid line). The total model consists of a non-accreting template star (blue dotted line) and three components with energy fluxes of $1 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$ (green dashed line), $1 \times 10^{11} \text{ erg s}^{-1} \text{ cm}^{-2}$ (grey dashed line) and $1 \times 10^{12} \text{ erg s}^{-1} \text{ cm}^{-2}$ (red dashed line). Best-fitting model parameters are listed in Extended Data Table 2. λ and F_λ are the wavelength and flux density, respectively.

Extended Data Fig. 3 Accretion flow modelling.

We show the best-fitting models (red dashed lines) to H α line profiles (black solid lines) measured with CHIRON and plotted in velocity, v . Best-fitting model parameters are listed in Extended Data Table 3. There is evidence for a wind that is absorbing part of the H α line profile on the blue side.

Extended Data Table 1 Swift observations and fluxes

Extended Data Table 2 Accretion-shock model parameters

Extended Data Table 3 Accretion-flow model parameters

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

Espaillat, C.C., Robinson, C.E., Romanova, M.M. *et al.* Measuring the density structure of an accretion hot spot. *Nature* **597**, 41–44 (2021).
<https://doi.org/10.1038/s41586-021-03751-5>

- Received: 04 March 2021
- Accepted: 21 June 2021
- Published: 01 September 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03751-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03751-5>

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

- Article
- [Published: 01 September 2021](#)

Emergent hydrodynamics in a strongly interacting dipolar spin ensemble

- [C. Zu](#)^{1,2} natl,
- [F. Machado](#) ORCID: orcid.org/0000-0003-0068-5073^{1,2} natl,
- [B. Ye](#) ORCID: orcid.org/0000-0002-1440-7843¹ natl,
- [S. Choi](#) ORCID: orcid.org/0000-0002-1247-062X¹,
- [B. Kobrin](#)^{1,2},
- [T. Mittiga](#)^{1,2},
- [S. Hsieh](#)^{1,2},
- [P. Bhattacharyya](#)^{1,2},
- [M. Markham](#)³,
- [D. Twitchen](#)³,
- [A. Jarmola](#)^{1,4},
- [D. Budker](#) ORCID: orcid.org/0000-0002-7356-4814^{1,5},
- [C. R. Laumann](#)⁶,
- [J. E. Moore](#)^{1,2} &
- [N. Y. Yao](#) ORCID: orcid.org/0000-0003-0194-7266^{1,2}

[Nature](#) volume **597**, pages 45–50 (2021)

- 1278 Accesses
- 23 Altmetric
- [Metrics details](#)

Subjects

- [Quantum simulation](#)
- [Statistical physics, thermodynamics and nonlinear dynamics](#)

Abstract

Conventional wisdom holds that macroscopic classical phenomena naturally emerge from microscopic quantum laws^{1,2,3,4,5,6,7}. However, despite this mantra, building direct connections between these two descriptions has remained an enduring scientific challenge. In particular, it is difficult to quantitatively predict the emergent ‘classical’ properties of a system (for example, diffusivity, viscosity and compressibility) from a generic microscopic quantum Hamiltonian^{7,8,9,10,11,12,13,14}. Here we introduce a hybrid solid-state spin platform, where the underlying disordered, dipolar quantum Hamiltonian gives rise to the emergence of unconventional spin diffusion at nanometre length scales. In particular, the combination of positional disorder and on-site random fields leads to diffusive dynamics that are Fickian yet non-Gaussian^{15,16,17,18,19,20}. Finally, by tuning the underlying parameters within the spin Hamiltonian via a combination of static and driven fields, we demonstrate direct control over the emergent spin diffusion coefficient. Our work enables the investigation of hydrodynamics in many-body quantum spin systems.

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: Nanoscale spin diffusion in a long-range interacting quantum system.



Fig. 2: Probing local spin-polarization dynamics using the NV centre.

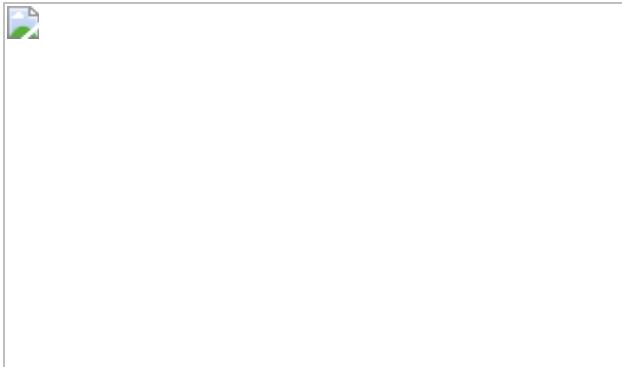
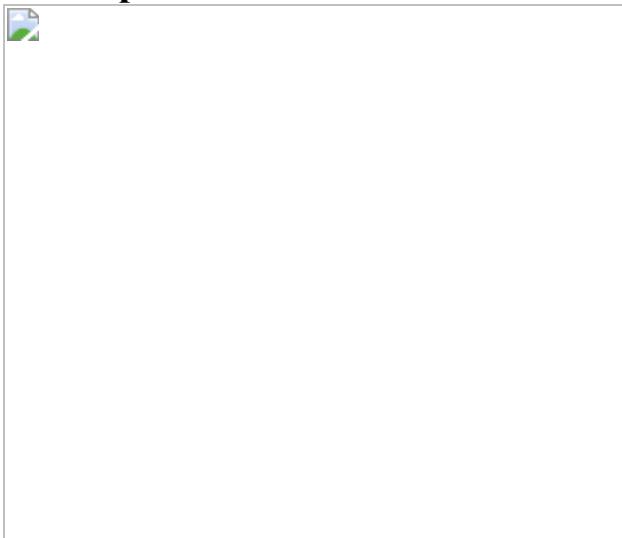


Fig. 3: Controlling emergent hydrodynamics by engineering the microscopic Hamiltonian.



Data availability

Source data are provided with this paper. Further data are available from the corresponding author upon reasonable request. [Source data](#) are provided with this paper.

References

1. 1.

- Ljubotina, M., Žnidarič, M. & Prosen, T. Spin diffusion from an inhomogeneous quench in an integrable system. *Nat. Commun.* **8**, 16117 (2017).

2. 2.

Sommer, A., Ku, M., Roati, G. & Zwierlein, M. W. Universal spin transport in a strongly interacting Fermi gas. *Nature* **472**, 201–204 (2011).

3. 3.

Moll, P. J., Kushwaha, P., Nandi, N., Schmidt, B. & Mackenzie, A. P. Evidence for hydrodynamic electron flow in PdCoO_2 . *Science* **351**, 1061–1064 (2016).

4. 4.

Bandurin, D. et al. Negative local resistance caused by viscous electron backflow in graphene. *Science* **351**, 1055–1058 (2016).

5. 5.

Crossno, J. et al. Observation of the dirac fluid and the breakdown of the Wiedemann–Franz law in graphene. *Science* **351**, 1058–1061 (2016).

6. 6.

Cepellotti, A. et al. Phonon hydrodynamics in two-dimensional materials. *Nat. Commun.* **6**, 6400 (2015).

7. 7.

Castro-Alvaredo, O. A., Doyon, B. & Yoshimura, T. Emergent hydrodynamics in integrable quantum systems out of equilibrium. *Phys. Rev. X* **6**, 041065 (2016).

8. 8.

Andreev, A., Kivelson, S. A. & Spivak, B. Hydrodynamic description of transport in strongly correlated electron systems. *Phys. Rev. Lett.* **106**, 256804 (2011).

9. 9.

Bertini, B., Collura, M., De Nardis, J. & Fagotti, M. Transport in out-of-equilibrium XXZ chains: exact profiles of charges and currents. *Phys. Rev. Lett.* **117**, 207201 (2016).

10. 10.

Bulchandani, V. B., Vasseur, R., Karrasch, C. & Moore, J. E. Solvable hydrodynamics of quantum integrable systems. *Phys. Rev. Lett.* **119**, 220604 (2017).

11. 11.

Ye, B., Machado, F., White, C. D., Mong, R. S. & Yao, N. Y. Emergent hydrodynamics in nonequilibrium quantum systems. *Phys. Rev. Lett.* **125**, 030601 (2020).

12. 12.

De Nardis, J., Bernard, D. & Doyon, B. Hydrodynamic diffusion in integrable systems. *Phys. Rev. Lett.* **121**, 160603 (2018).

13. 13.

Bertini, B. et al. Finite-temperature transport in one-dimensional quantum lattice models. *Rev. Mod. Phys.* **93**, 025003 (2021).

14. 14.

Ku, M. J. et al. Imaging viscous flow of the Dirac fluid in graphene. *Nature* **583**, 537–541 (2020).

15. 15.

Chechkin, A. V., Seno, F., Metzler, R. & Sokolov, I. M. Brownian yet non-Gaussian diffusion: from superstatistics to subordination of diffusing diffusivities. *Phys. Rev. X* **7**, 021002 (2017).

16. 16.

Stylianidou, S., Kuwada, N. J. & Wiggins, P. A. Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility. *Biophys. J.* **107**, 2684–2692 (2014).

17. 17.

Kim, J., Kim, C. & Sung, B. J. Simulation study of seemingly Fickian but heterogeneous dynamics of two dimensional colloids. *Phys. Rev. Lett.* **110**, 047801 (2013).

18. 18.

Chubynsky, M. V. & Slater, G. W. Diffusing diffusivity: a model for anomalous, yet brownian, diffusion. *Phys. Rev. Lett.* **113**, 098302 (2014).

19. 19.

Postnikov, E. B., Chechkin, A. & Sokolov, I. M. Brownian yet non-Gaussian diffusion in heterogeneous media: from superstatistics to homogenization. *New J. Phys.* **22**, 063046 (2020).

20. 20.

Barkai, E. & Burov, S. Packets of diffusing particles exhibit universal exponential tails. *Phys. Rev. Lett.* **124**, 060603 (2020).

21. 21.

Wurtz, J. & Polkovnikov, A. Quantum hydrodynamics in spin chains with phase space methods. *Phys. Rev. E* **101**, 052120 (2020).

22. 22.

Khemani, V., Vishwanath, A. & Huse, D. A. Operator spreading and the emergence of dissipative hydrodynamics under unitary evolution with conservation laws. *Phys. Rev. X* **8**, 031057 (2018).

23. 23.

Levstein, P., Pastawski, H. & Calvo, R. Spin diffusion in low-dimensional copper-amino-acid complexes. *J. Phys. Condens. Matter* **3**, 1877–1888 (1991).

24. 24.

Narozhny, B., Gornyi, I., Titov, M., Schütt, M. & Mirlin, A. Hydrodynamics in graphene: linear-response transport. *Phys. Rev. B* **91**, 035414 (2015).

25. 25.

Moore, G. D. & Sohrabi, K. A. Kubo formulas for second-order hydrodynamic coefficients. *Phys. Rev. Lett.* **106**, 122302 (2011).

26. 26.

Friedman, A. J., Gopalakrishnan, S. & Vasseur, R. Diffusive hydrodynamics from integrability breaking. *Phys. Rev. B* **101**, 180302 (2020).

27. 27.

Schemmer, M., Bouchoule, I., Doyon, B. & Dubail, J. Generalized hydrodynamics on an atom chip. *Phys. Rev. Lett.* **122**, 090601 (2019).

28. 28.

Zhang, W. & Cory, D. First direct measurement of the spin diffusion rate in a homogenous solid. *Phys. Rev. Lett.* **80**, 1324–1327 (1998).

29. 29.

Pagliero, D. et al. Optically pumped spin polarization as a probe of many-body thermalization. *Sci. Adv.* **6**, eaaz6986 (2020).

30. 30.

Boutis, G., Greenbaum, D., Cho, H., Cory, D. & Ramanathan, C. Spin diffusion of correlated two-spin states in a dielectric crystal. *Phys. Rev. Lett.* **92**, 137201 (2004).

31. 31.

Eberhardt, K. W., Mouaziz, S., Boero, G., Brugger, J. & Meier, B. H. Direct observation of nuclear spin diffusion in real space. *Phys. Rev. Lett.* **99**, 227603 (2007).

32. 32.

Kardar, M., Parisi, G. & Zhang, Y.-C. Dynamic scaling of growing interfaces. *Phys. Rev. Lett.* **56**, 889–892 (1986).

33. 33.

Žnidarič, M., Scardicchio, A. & Varma, V. K. Diffusive and subdiffusive spin transport in the ergodic phase of a many-body localizable system. *Phys. Rev. Lett.* **117**, 040601 (2016).

34. 34.

Gopalakrishnan, S. & Vasseur, R. Kinetic theory of spin diffusion and superdiffusion in XXZ spin chains. *Phys. Rev. Lett.* **122**, 127202 (2019).

35. 35.

Sinai, Y. G. The limiting behavior of a one-dimensional random walk in a random medium. *Theory Probab. Appl.* **27**, 256–268 (1983).

36. 36.

Doherty, M. W. et al. The nitrogen-vacancy colour centre in diamond. *Phys. Rep.* **528**, 1–45 (2013).

37. 37.

Hall, L. et al. Detection of nanoscale electron spin resonance spectra demonstrated using nitrogen-vacancy centre probes in diamond. *Nat. Commun.* **7**, 10211 (2016).

38. 38.

Khutsishvili, G. Spin diffusion and magnetic relaxation of nuclei. *Soviet Phys. JETP* **15**, (1962).

39. 39.

Takahashi, S., Hanson, R., Van Tol, J., Sherwin, M. S. & Awschalom, D. D. Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.* **101**, 047601 (2008).

40. 40.

Hunt, G. A. Some theorems concerning brownian motion. *Trans. Am. Math. Soc.* **81**, 294–319 (1956).

41. 41.

Jarmola, A., Acosta, V., Jensen, K., Chemerisov, S. & Budker, D. Temperature-and magnetic-field-dependent longitudinal spin relaxation in nitrogen-vacancy ensembles in diamond. *Phys. Rev. Lett.* **108**, 197601 (2012).

42. 42.

Spitzer, F. *Principles of Random Walk* Vol. 34 (Springer Science & Business Media, 2013).

43. 43.

Einstein, A. *Investigations on the Theory of the Brownian Movement* (Courier Corporation, 1956).

44. 44.

Lévy, P. *Théorie de l'Addition des Variables Aléatoires* Vol. 1 (Gauthier-Villars, 1954).

45. 45.

Abanin, D. A., Altman, E., Bloch, I. & Serbyn, M. Colloquium: many-body localization, thermalization, and entanglement. *Rev. Mod. Phys.* **91**, 021001 (2019).

46. 46.

Mori, T., Ikeda, T. N., Kaminishi, E. & Ueda, M. Thermalization and prethermalization in isolated quantum systems: a theoretical overview. *J. Phys. B* **51**, 112001 (2018).

47. 47.

Yao, N. Y. et al. Many-body localization in dipolar systems. *Phys. Rev. Lett.* **113**, 243002 (2014).

48. 48.

Nandkishore, R. M. & Sondhi, S. L. Many-body localization with long-range interactions. *Phys. Rev. X* **7**, 041021 (2017).

49. 49.

Peng, P., Yin, C., Huang, X., Ramanathan, C. & Cappellaro, P. Observation of floquet prethermalization in dipolar spin chains. *Nature Physics* **17**, 444–447 (2021).

50. 50.

Rittweger, E., Han, K. Y., Irvine, S. E., Eggeling, C. & Hell, S. W. STED microscopy reveals crystal colour centres with nanometric resolution. *Nat. Photon.* **3**, 144–147 (2009).

51. 51.

Arai, K. et al. Fourier magnetic imaging with nanoscale resolution and compressed sensing speed-up using electronic spins in diamond. *Nat. Nanotechnol.* **10**, 859–864 (2015).

52. 52.

Choi, J. et al. Robust dynamic hamiltonian engineering of many-body spin systems. *Phys. Rev. X* **10**, 031002 (2020).

53. 53.

Zhou, H. et al. Quantum metrology with strongly interacting spin systems. *Phys. Rev. X* **10**, 031003 (2020).

54. 54.

Yukawa, H. On the interaction of elementary particles. i. *Proc. Phys. Math. Soc. Jpn* **17**, 48–57 (1935).

55. 55.

Shlesinger, M. F., Zaslavsky, G. M. & Frisch, U. (eds) *Lévy Flights and Related Topics in Physics: Proc. of the International Workshop Held at Nice, France, 27–30 June 1994* (Lecture Notes in Physics, Springer-Verlag, 1995).

56. 56.

Schuckert, A., Lovas, I. & Knap, M. Nonlocal emergent hydrodynamics in a long-range quantum spin system. *Phys. Rev. B* **101**, 020416 (2020).

57. 57.

Bauch, E. et al. Decoherence of ensembles of nitrogen-vacancy centers in diamond. *Phys. Rev. B* **102**, 134210 (2020).

58. 58.

Mittiga, T. et al. Imaging the local charge environment of nitrogen-vacancy centers in diamond. *Phys. Rev. Lett.* **121**, 246402 (2018).

Acknowledgements

We acknowledge the insights of and discussions with J. Choi, E. Davis, M. Dupont, D. Gangloff, S. Gopalakrishnan, A. Jayich, P. Stamp, R. Walsworth and H. Zhou. This work was supported as part of the Center for Novel Pathways to Quantum Coherence in Materials, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under award number DE-AC02-05CH11231. A.J. acknowledges support from the Army Research Laboratory under Cooperative Agreement number W911NF-16-2-0008. S.H. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant number DGE1752814. N.Y.Y. acknowledges support from the David and Lucile Packard foundation and the W. M. Keck foundation. The work of D.B. is supported by the EU FET-OPEN Flagship Project ASTERIQS (action 820394 and the Cluster of Excellence ‘Precision Physics, Fundamental Interactions, and Structure of Matter’ (PRISMA+ EXC 2118/1) funded by the German Research Foundation (DFG) within the German Excellence Strategy (project ID 39083149). C.R.L. acknowledges support from the NSF through PHY-1752727. S.C. acknowledges support from the Miller Institute for Basic Research in Science.

Author information

Author notes

1. These authors contributed equally: C. Zu, F. Machado, B. Ye

Affiliations

1. Department of Physics, University of California, Berkeley, CA, USA
C. Zu, F. Machado, B. Ye, S. Choi, B. Kobrin, T. Mittiga, S. Hsieh, P. Bhattacharyya, A. Jarmola, D. Budker, J. E. Moore & N. Y. Yao

2. Materials Science Division, Lawrence Berkeley National Laboratory,
Berkeley, CA, USA

C. Zu, F. Machado, B. Kobrin, T. Mittiga, S. Hsieh, P.
Bhattacharyya, J. E. Moore & N. Y. Yao

3. Element Six, Harwell, UK

M. Markham & D. Twitchen

4. US Army Research Laboratory, Adelphi, MD, USA

A. Jarmola

5. Helmholtz Institut Mainz, Johannes Gutenberg Universitat Mainz,
Mainz, Germany

D. Budker

6. Department of Physics, Boston University, Boston, MA, USA

C. R. Laumann

Contributions

C.Z., T.M., S.H. and P.B. performed the experiments. F.M., B.Y., S.C.,
B.K., C.R.L., J.E.M. and N.Y.Y. developed the theoretical models and
methodology. C.Z., F.M., and B.Y. performed the data analysis. F.M. and
B.Y. performed the numerical simulations. M.M., D.T., A.J. and D.B.
prepared and provided the diamond substrates. C.R.L., J.E.M. and N.Y.Y.
supervised the project. C.Z., F.M., B.Y., C.R.L., J.E.M. and N.Y.Y wrote the
manuscript with input from all authors.

Corresponding author

Correspondence to [N. Y. Yao](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Nir Bar-Gill, Benjamin Doyon 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 Experimental sequence to measure the coherence of P1 ensemble.](#)

For XY-8 and interaction decoupling sequences, we fix the interval between pulses to be $\tau = 10$ ns, and increase the number of repetition N . The P1 $\pi(\pi/2)$ pulse duration is set to 36 ns (18 ns). In the interaction decoupling sequences, the pulses at bottom side correspond to rotations along $(-\hat{x})$ (blue) and $(-\hat{y})$ (orange) axes.

[Extended Data Fig. 2 Extraction of diffusion constants for sample S1.](#)

Fitting of the depolarization data in sample S1 for different groups v (different panels) and different pumping times τ_p (different colours). For each group, we fit the experimental data across all τ_p data to a diffusive model in equation (7) with an additional background P_{bg} . From this procedure, we extract both D and b , as well as, a τ_p -dependent Γ , which captures the reduction in efficiency of the NV–P1 polarization transfer owing to the saturation of polarization near the NV

[Source data](#).

Extended Data Fig. 3 Measurement of late time extrinsic decay time T_1 for different samples and under driving.

a, b, Extraction of the extrinsic depolarization time of samples S1 and S2 at room and low temperature (25 K) and after polarizing for $\tau_p = 1,000 \mu\text{s}$ (**a**) and $\tau_p = 30 \mu\text{s}$ (**b**). The late time behaviour follows an exponential decay with timescale given by $1.0 \pm 0.1 \text{ ms}$ and $2.6 \pm 0.2 \text{ ms}$, respectively. **c**, To extract the modified intrinsic depolarization time $\langle\langle T_1 \rangle\rangle$ of $v = 1/4$ P1 subgroup with the presence of a strong microwave driving $\Omega = (2\pi) \times 11.7 \text{ MHz}$ on the other $v = 1/4$ subgroup, we apply the following pulse sequence: after a laser pumping time $\tau_p = 1000 \mu\text{s}$, we wait for 1 ms so that the initial P1 spatial polarization profile diffuses to a nearly homogeneous background which decays with intrinsic depolarization time of P1 centres. We then turn on a continuous microwave driving on the other $v = 1/4$ P1 subgroup, and measure the resulting background decay; the resulting timescale is given $\langle\langle T_1 \rangle\rangle = 0.9 \pm 0.2 \text{ ms}$

[Source data](#).

Extended Data Fig. 4 Determination of the extrinsic decoherence rate γ and on-site random field distribution δ_i .

a, We estimate the extrinsic decoherence rate γ in the rate equation using the measured spin-echo coherence time of the NV. After polarizing the NV centre via a green laser, a $\pi/2$ pulse prepares the NV spin into a coherent superposition of $|m=0\rangle$ and $|m=-1\rangle$, which is allowed to dephase during a time t . A π pulse at the centre of the sequence ‘echos’ out the on-site random field generated by the nearby P1 centres, thus provides a direct estimation of the extrinsic decoherence time of a single spin in the system. We fit the spin-echo decay using a form $\langle\langle \rm{e} \rangle\rangle^{-\frac{(t-T_2)^2}{T_2^2+1.5}}$ (ref. [57](#)) and extract $\langle\langle T_2 \rangle\rangle = 1.9 \pm 0.1 \text{ ms}$ and $\langle\langle \gamma \rangle\rangle \approx 1/\langle\langle T_2 \rangle\rangle \approx 0.5 \text{ ms}^{-1}$

$\{\text{rm}\{s\}\}^{\{-1\}}\}$). **b**, The distribution of on-site random fields δ_i is directly determined using the intrinsic linewidth of the NV spin state. After polarizing the NV centre via a green laser, we apply a microwave π pulse and sweep its frequency ω across the NV $|0\rangle$ to $| -1 \rangle$ transition. To avoid microwave power broadening of NV transition, we choose a sufficient weak microwave π pulse with duration 2 μs . Note that the measured linewidth is dominated by interactions with the dense P1 ensemble ($W \approx (2\pi) \times 4.5 \text{ MHz}$) ([Supplementary Information](#)). The presence of nuclear ^{13}C spins leads to a much smaller contribution to the linewidth of about $(2\pi) \times 0.3 \text{ MHz}$ (ref. [58](#)). Crucially, both effects are taken into account in our analysis by sampling δ_i directly from the measured spectrum

[Source data](#).

[**Extended Data Fig. 5 Agreement between semiclassical model and experimentally observed dynamics.**](#)

Given the approximately equal P1 density of both the sample S1 and S2, we simulate the the dynamics of a single NV defect surrounded by $N_{\text{P1s}} = \{300, 225, 75\}$ P1 centres for the groups $v = \{1/3, 1/4, 1/12\}$, respectively. In the polarization protocol, we choose $\Gamma_p = 0.1 \mu\text{s}^{-1}$ for S1 and $\Gamma_p = 0.25 \mu\text{s}^{-1}$ for S2. The subsequent polarization dynamics of the NV centre is given by the difference in populations between the $|0\rangle$ and $| -1 \rangle$ states. For the $v \in \{1/3, 1/4\}$ groups of S1, we observe excellent agreement with the experimental data for over four orders of magnitude in τ_p and throughout then entire experimental timescale using $\gamma = 0.5 \mu\text{s}^{-1}$. For the $v = 1/12$ group of S1, we observe good agreement, albeit with a smaller range of τ_p and using $\gamma = 1.5 \mu\text{s}^{-1}$. We believe this discrepancy arises from a much larger separation between the strength of the on-site fields and the flip-flop rate of the ensemble. For the $v = 1/3$ group of S2, we also observe excellent agreement throughout the entire dynamics, where we use $\gamma = 0.3 \mu\text{s}^{-1}$. The agreement observed in the NV polarization decay in both samples gives us confidence that our semiclassical model can capture the polarization

dynamics in the sample and provide an accurate calculation of the diffusive properties of the spin ensemble

[Source data](#).

Extended Data Fig. 6 Summary of extraction of diffusion coefficient.

a–c, Extraction of diffusion coefficient of sample S2 at low temperature. **a**, Growth of $\langle \langle r \rangle^2 \rangle$ for different system sizes N and the infinite system scaling (black line). **b**, Finite size scaling of $\langle \langle r \rangle^2 \rangle$ to $N \rightarrow \infty$ assuming a linear in $L^{-1} \approx N^{-1/3}$ correction for representative values of t . **c**, Fitting the early-time growth of $\langle \langle r \rangle^2 \rangle$ up to different times $T_{\max} \in [30, 300]$ leads to slightly different values of the diffusion coefficient, whether including a constant offset (light blue) or not (dark blue). Considering the fit without an offset, the final diffusion coefficient is taken to be the average with an uncertainty given by half the range of diffusion coefficients. **d**, For the different experimental conditions using the parameters discussed in Methods, we extract the diffusion coefficient from the growth of $\langle \langle r \rangle^2 \rangle$, which is in great agreement with the experimentally extracted values after correcting for the non-Gaussian polarization profile (Table 1)

[Source data](#).

Extended Data Fig. 7 Determination of the length scale $\langle \langle \ell \rangle \rangle$.

Extracted $\langle \langle \ell \rangle \rangle$ for different samples and different P1 groups as a function of the early time cut-off t_{\min} . Averaging over the last three data points, where the $\langle \langle \ell \rangle \rangle$ are consistent, yields the reported value of $\langle \langle \ell \rangle \rangle$. The red dashed line corresponds to the final value and the shaded area is the associated uncertainty

[Source data](#).

Extended Data Fig. 8 Long-range modification to conventional diffusion.

The presence of a long-range k^3 -term parametrically modifies the approach, $\langle A \rangle - \langle \langle \rm{rm}\{p\} \rangle \rangle(t) = \langle S \rangle - \langle \langle \rm{rm}\{p\} \rangle \rangle(t) - \langle (4 \langle \rm{rm}\{\pi\} \rangle D t) \rangle^{-3/2}$, to the late-time Gaussian fixed point, as highlighted in a three-dimensional, disorder-less numerical simulation, with lattice constant a and diffusion coefficient D

[Source data](#).

Extended Data Fig. 9 Fitting of experimental data with different modifications to diffusion equation.

Fitting of the diffusive description with different terms and fixed $T_1 = 2.6$ ms in sample S2 with $\tau_p = 30 \mu\text{s}$. Different columns represent fitting to a different range of the data (highlighted by the red shaded region). The inclusion of more terms in the diffusive description allows for a better fit of the data; however, the improvement in the fitting range is only significant when the fitting regimes includes early time data ($\leq 30 \mu\text{s}$), as highlighted in the second row of the relative residuals. All data are presented with logarithmically spaced y axis, except in the grey shaded region where a linear regime is used to highlight the fluctuations of the residuals around 0. Fits in Fig. 1b correspond to the third column

[Source data](#).

Extended Data Table 1 Extracted $\langle \langle \ell \rangle \rangle$ from the spin polarization dynamics for the different sample considered (S1 and S2) and the different P1 subgroups

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Discussion, including Supplementary Figs. 1–10, and additional references.

[Peer Review File](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

[Source Data Extended Data Fig. 8](#)

[Source Data Extended Data Fig. 9](#)

[Source Data Extended Data Table 1](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Zu, C., Machado, F., Ye, B. *et al.* Emergent hydrodynamics in a strongly interacting dipolar spin ensemble. *Nature* **597**, 45–50 (2021).
<https://doi.org/10.1038/s41586-021-03763-1>

- Received: 22 September 2020
- Accepted: 23 June 2021
- Published: 01 September 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03763-1>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03763-1>

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

- Article
- [Published: 01 September 2021](#)

Decision trees within a molecular memristor

- [Sreetosh Goswami](#) ORCID: [orcid.org/0000-0002-5104-5757](#)^{1,2,3},
- [Rajib Pramanick](#) ORCID: [orcid.org/0000-0001-8876-7238](#)⁴ na1,
- [Abhijeet Patra](#) ORCID: [orcid.org/0000-0003-0422-3319](#)^{1,5} na1,
- [Santi Prasad Rath](#) ORCID: [orcid.org/0000-0003-0333-9993](#)⁴ na1,
- [Martin Foltin](#)⁶,
- [A. Ariando](#) ORCID: [orcid.org/0000-0002-0598-426X](#)^{1,2,3},
- [Damien Thompson](#) ORCID: [orcid.org/0000-0003-2340-5441](#)⁷,
- [T. Venkatesan](#) ORCID: [orcid.org/0000-0001-9683-4584](#)^{1,2,3,8,9},
- [Sreebrata Goswami](#) ORCID: [orcid.org/0000-0002-4380-5656](#)⁴ &
- [R. Stanley Williams](#) ORCID: [orcid.org/0000-0003-0213-4259](#)¹⁰

[Nature](#) volume 597, pages 51–56 (2021)

- 3831 Accesses
- 377 Altmetric
- [Metrics details](#)

Subjects

- [Electrical and electronic engineering](#)
- [Electronic and spintronic devices](#)
- [Electronic devices](#)

Abstract

Profuse dendritic-synaptic interconnections among neurons in the neocortex embed intricate logic structures enabling sophisticated decision-making that vastly outperforms any artificial electronic analogues^{1,2,3}. The physical complexity is far beyond existing circuit fabrication technologies: moreover, the network in a brain is dynamically reconfigurable, which provides flexibility and adaptability to changing environments^{4,5,6}. In contrast, state-of-the-art semiconductor logic circuits are based on threshold switches that are hard-wired to perform predefined logic functions. To advance the performance of logic circuits, we are re-imagining fundamental electronic circuit elements by expressing complex logic in nanometre-scale material properties. Here we use voltage-driven conditional logic interconnectivity among five distinct molecular redox states of a metal-organic complex to embed a ‘thicket’ of decision trees (composed of multiple if-then-else conditional statements) having 71 nodes within a single memristor. The resultant current–voltage characteristic of this molecular memristor (a ‘memory resistor’, a globally passive resistive-switch circuit element that axiomatically complements the set of capacitor, inductor and resistor) exhibits eight recurrent and history-dependent non-volatile switching transitions between two conductance levels in a single sweep cycle. The identity of each molecular redox state was determined with *in situ* Raman spectroscopy and confirmed by quantum chemical calculations, revealing the electron transport mechanism. Using simple circuits of only these elements, we experimentally demonstrate dynamically reconfigurable, commutative and non-commutative stateful logic in multivariable decision trees that execute in a single time step and can, for example, be applied as local intelligence in edge computing^{7,8,9}.

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: Circuit element structure and $I(V)$ characteristics.

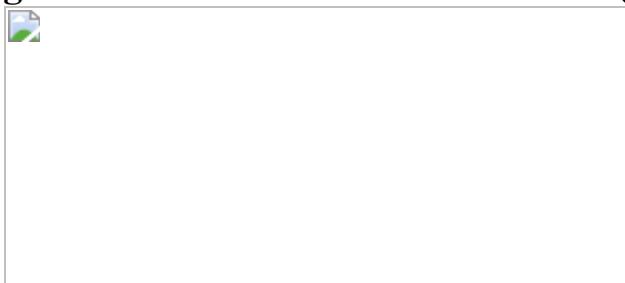


Fig. 2: In situ spectroscopy.

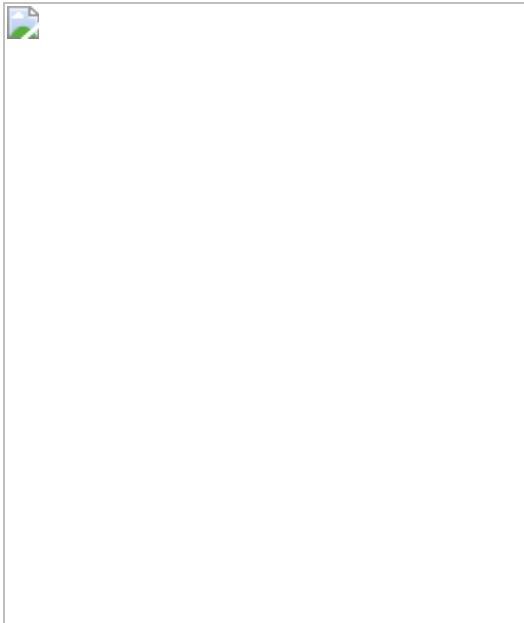


Fig. 3: Decision thicket within a single molecular film circuit element.

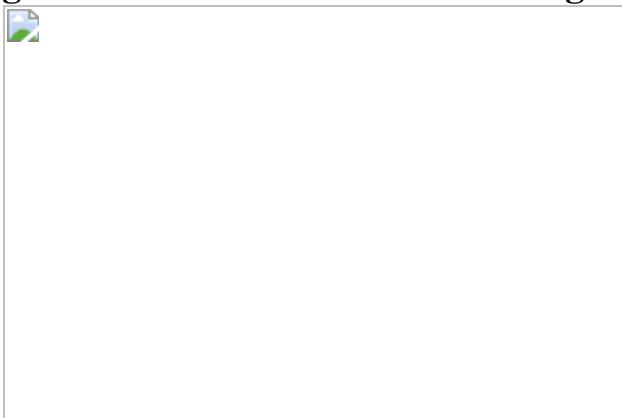
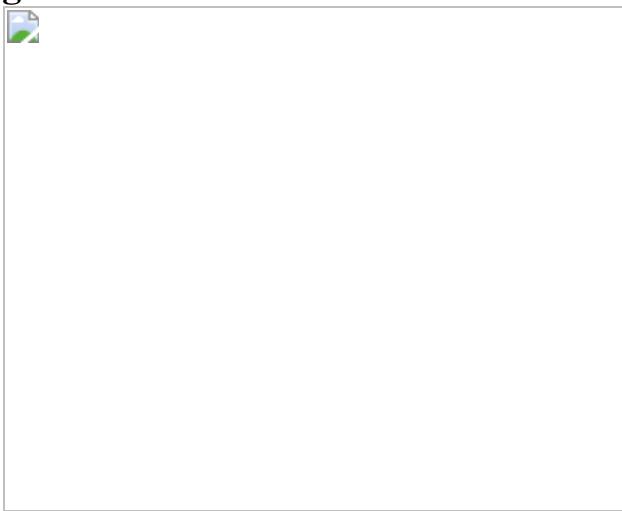


Fig. 4: Realization of multivariable decision trees.



Data availability

The data on which the figures are constructed and from which the conclusions are drawn are available from the corresponding authors upon reasonable request.

References

1. 1.

Park, H.-J. & Friston, K. Structural and functional brain networks: from connections to cognition. *Science* **342**, 1238411 (2013).

2. 2.

Eliasmith, C. et al. A large-scale model of the functioning brain. *Science* **338**, 1202–1205 (2012).

3. 3.

Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).

4. 4.

Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).

5. 5.

Braun, U. et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl Acad. Sci. USA* **112**, 11678–11683 (2015).

6. 6.

Huang, Z., Zhang, J., Wu, J., Mashour, G. A. & Hudetz, A. G. Temporal circuit of macroscale dynamic brain activity supports human

consciousness. *Sci. Adv.* **6**, eaaz0087 (2020).

7. 7.

Take it to the edge. *Nat. Electron.* **2**, 1 (2019).

8. 8.

Vaughan, O. Working on the edge. *Nat. Electron.* **2**, 2–3 (2019).

9. 9.

Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).

10. 10.

Shannon, C. *A Symbolic Analysis of Relay and Switching Circuits*. MS thesis, MIT (1940).

11. 11.

Goswami, S. et al. Robust resistive memory devices using solution-processable metal-coordinated azo aromatics. *Nat. Mater.* **16**, 1216–1224 (2017); erratum **17**, 103 (2018)0.

12. 12.

Goswami, S., Thompson, D., Williams, R. S., Goswami, S. & Venkatesan, T. Colossal current and voltage tunability in an organic memristor via electrode engineering. *Appl. Mater. Today* **19**, 100626 (2020).

13. 13.

Goswami, S. et al. Charge disproportionate molecular redox for discrete memristive and memcapacitive switching. *Nat. Nanotechnol.* **15**, 380–389 (2020).

14. 14.

Goswami, S., Goswami, S. & Venkatesan, T. An organic approach to low energy memory and brain inspired electronics. *Appl. Phys. Rev.* **7**, 021303 (2020).

15. 15.

Goswami, S. et al. Nanometer-scale uniform conductance switching in molecular memristors. *Adv. Mater.* **32**, 2004370 (2020).

16. 16.

Goswami, S. *Resistive Memories Using Metal-Azo Aromatics*. Ph.D. thesis, National Univ. Singapore (2018).

17. 17.

Baldwin, D. A., Lever, A. B. & Parish, R. V. Complexes of 2,2'-azopyridine with iron(II), cobalt(II), nickel(II), copper(I), and copper(II). Infrared study. *Inorg. Chem.* **8**, 107–115 (1969).

18. 18.

Nagai, K. & Kitagawa, T. Differences in Fe(II)-N epsilon (His-F8) stretching frequencies between deoxyhemoglobins in the two alternative quaternary structures. *Proc. Natl Acad. Sci. USA* **77**, 2033–2037 (1980).

19. 19.

Benko, B. & Yu, N.-T. Resonance Raman studies of nitric oxide binding to ferric and ferrous hemoproteins: detection of Fe(III)–NO stretching, Fe(III)–N–O bending, and Fe(II)–N–O bending vibrations. *Proc. Natl Acad. Sci. USA* **80**, 7042–7046 (1983).

20. 20.

Miller, J. S. & Min, K. S. Oxidation leading to reduction: redox-induced electron transfer (RIET). *Angew. Chem. Int. Edn* **48**, 262–272 (2009).

21. 21.

Sengupta, D. et al. Size-selective Pt siderophores based on redox-active azo-aromatic ligands. *Chem. Sci.* **11**, 9226–9236 (2020).

22. 22.

Gass, I. A. et al. Anion dependent redox changes in iron bis-terdentate nitroxide {NNO} chelates. *Inorg. Chem.* **50**, 3052–3064 (2011).

23. 23.

Winkler, C. et al. Understanding the correlation between electronic coupling and energetic stability of molecular crystal polymorphs: the instructive case of quinacridone. *Chem. Mater.* **31**, 7054–7069 (2019).

24. 24.

Baldoni, M., Lorenzoni, A., Pecchia, A. & Mercuri, F. Spatial and orientational dependence of electron transfer parameters in aggregates of iridium-containing host materials for OLEDs: coupling constrained density functional theory with molecular dynamics. *Phys. Chem. Chem. Phys.* **20**, 28393–28399 (2018).

25. 25.

Knuth, D. E. *The Art of Computer Programming* Vol. 3 (Pearson Education, 1997).

26. 26.

Loh, W. Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **1**, 14–23 (2011).

27. 27.

Altman, N. & Krzywinski, M. Ensemble methods: bagging and random forests. *Nat. Methods* **14**, 933–934 (2017).

28. 28.

Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16**, 321–332 (2015).

29. 29.

Choi, M. Y. & Ma, C. Making a big impact with small datasets using machine-learning approaches. *Lancet Rheumatol.* **2**, e451–e452 (2020).

30. 30.

Rani, P., Sarkar, N. & Liu, C. Maintaining optimal challenge in computer games through real-time physiological feedback. In *Foundations of Augmented Cognition* (ed. Schmorow, D. D.) 184 (Taylor & Francis, 2005).

31. 31.

Borghetti, J. et al. ‘Memristive’ switches enable ‘stateful’ logic operations via material implication. *Nature* **464**, 873–876 (2010).

32. 32.

Wang, Z. et al. Resistive switching materials for information processing. *Nat. Rev. Mater.* **5**, 173–195 (2020).

33. 33.

Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544 (2020); correction **15**, 812 (2020).

34. 34.

Gidon, A. et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* **367**, 83–87 (2020).

35. 35.

Kim, K. M. & Williams, R. S. A family of stateful memristor gates for complete cascading logic. *IEEE Trans. Circuits Syst. I* **66**, 4348–4355 (2019).

36. 36.

Kim, Y. S. et al. Stateful in-memory logic system and its practical implementation in a TaO_x-based bipolar-type memristive crossbar array. *Adv. Intell. Syst.* **2**, 1900156 (2020).

37. 37.

Shen, W. et al. Stateful logic operations in one-transistor-one-resistor resistive random access memory array. *IEEE Electron Device Lett.* **40**, 1538–1541 (2019).

38. 38.

Xu, N. et al. A stateful logic family based on a new logic primitive circuit composed of two antiparallel bipolar memristors. *Adv. Intell. Syst.* **2**, 1900082 (2020).

39. 39.

Li, C. et al. Analog content-addressable memories with memristors. *Nat. Commun.* **11**, 1–8 (2020).

40. 40.

Kim, Y. S., Son, M. W. & Kim, K. M. Memristive stateful logic for edge Boolean computers. *Adv. Intell. Syst.* **3**, 2000278 (2021).

41. 41.

Du Nguyen, H. A. et al. in *Proc. 2015 IEEE/ACM Int. Symp. on Nanoscale Architectures (NANOARCH' 15)* 57–62 (IEEE).

42. 42.

Du Nguyen, H. A. et al. On the implementation of computation-in-memory parallel adder. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **25**, 2206–2219 (2017).

43. 43.

Yu, J., Lebdeh, M. A., Du Nguyen, H. A., Taouil, M. & Hamdioui, S. in *Proc. 2020 25th Asia and South Pacific Design Automation Conf. (ASP-DAC)* 385–392 (IEEE) (2020).

44. 44.

Du Nguyen, H. A. et al. in *Proc. 2017 IFIP/IEEE Int. Conf. on Very Large Scale Integration (VLSI-SoC)* 1–10 (IEEE) (2017).

Acknowledgements

This research was supported by the Singapore National Research Foundation (NRF) under the Competitive Research Programs (CRP award no. NRF-CRP15-2015-01). Sreebrata Goswami acknowledges the financial support of SERB, India, through grants SR/S2/JCB-09/2011. R.S.W. acknowledges the X-Grants Program of the President's Excellence Fund at Texas A&M University. A.A. thanks the Agency for Science, Technology and Research (A*STAR) for its Advanced Manufacturing and Engineering (AME) Individual Research Grant (IRG), A1983c0034. D.T. acknowledges support from Science Foundation Ireland (SFI) under awards 15/CDA/3491 and 12/RC/2275_P2, and supercomputing resources at the SFI/Higher Education Authority Irish Center for High-End Computing (ICHEC). We thank S. Hooda for RBS data, L. Haidong for making the shadow masks, S. Sarkar for helping with the in situ spectroscopic measurement setup, A. Tampèz for tip-induced current and spectroscopic measurements, F. Meyer and S. Demeshko for Mössbauer measurements and D. Deb for inputs

regarding the figures. We thank W. Robinett, C. R. Sinha and Q. Xia for their comments and suggestions.

Author information

Author notes

1. These authors contributed equally: Rajib Pramanick, Abhijeet Patra, Santi Prasad Rath

Affiliations

1. Department of Physics, National University of Singapore, Singapore, Singapore

Sreetosh Goswami, Abhijeet Patra, A. Ariando & T. Venkatesan

2. NUSNNI-NanoCore, National University of Singapore, Singapore, Singapore

Sreetosh Goswami, A. Ariando & T. Venkatesan

3. NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore

Sreetosh Goswami, A. Ariando & T. Venkatesan

4. School of Chemical Sciences, Indian Association for the Cultivation of Science (IACS), Kolkata, India

Rajib Pramanick, Santi Prasad Rath & Sreebrata Goswami

5. On Deck, San Francisco, CA, USA

Abhijeet Patra

6. AI Research Lab, Hewlett Packard Enterprise, Fort Collins, CO, USA

Martin Foltin

7. Department of Physics, Bernal Institute, University of Limerick,
Limerick, Ireland

Damien Thompson

8. Department of Electrical and Computer Engineering, National
University of Singapore, Singapore, Singapore

T. Venkatesan

9. Center for Quantum Research and Technology, The University of
Oklahoma, Norman, OK, USA

T. Venkatesan

10. Department of Electrical and Computer Engineering, Texas A&M
University, College Station, TX, USA

R. Stanley Williams

Contributions

Sreetosh Goswami devised the project and performed the electrical and spectroscopic measurements. R.S.W. and Sreetosh Goswami designed the experiments and performed data analysis. Sreebrata Goswami introduced the materials and proposed the molecular mechanism. R.P. and S.P.R. performed molecular synthesis and characterization. M.F. provided insights into circuit design. D.T. conceived and performed the electronic structure calculations. Sreetosh Goswami, T.V., D.T., M.F., A.P., A.A., Sreebrata Goswami and R.S.W. discussed and wrote the paper.

Corresponding authors

Correspondence to [Sreetosh Goswami](#) or [T. Venkatesan](#) or [Sreebrata Goswami](#) or [R. Stanley Williams](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Matthew Marinella, Ilia Valov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary Information

This document contains 9 sections of Supplementary Discussion, with 47 Supplementary Figures and 4 Supplementary Tables providing additional information on circuit element fabrication, material characterisation, electrical, spectroscopic measurements, and data analysis supporting the main text.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Goswami, S., Pramanick, R., Patra, A. *et al.* Decision trees within a molecular memristor. *Nature* **597**, 51–56 (2021).
<https://doi.org/10.1038/s41586-021-03748-0>

- Received: 05 November 2020
- Accepted: 21 June 2021
- Published: 01 September 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03748-0>

Molecular memristors offer a path to ultra-efficient computing

- Matthew J. Marinella
- A. Alec Talin

News & Views 01 Sept 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03748-0>

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

- Article
- [Published: 01 September 2021](#)

Scalable production of high-performing woven lithium-ion fibre batteries

- [Jiqing He^{1,2,3 na1}](#),
- [Chenhao Lu^{1,2,3 na1}](#),
- [Haibo Jiang^{1,2,3}](#),
- [Fei Han](#) ORCID: [orcid.org/0000-0002-5143-9836⁴](https://orcid.org/0000-0002-5143-9836),
- [Xiang Shi^{1,2,3}](#),
- [Jingxia Wu^{1,2,3}](#),
- [Liyuan Wang^{1,2,3}](#),
- [Taiqiang Chen](#) ORCID: [orcid.org/0000-0002-4018-2561^{1,2,3}](https://orcid.org/0000-0002-4018-2561),
- [Jiajia Wang^{1,2,3}](#),
- [Ye Zhang^{1,2,3}](#),
- [Han Yang^{1,2,3}](#),
- [Guoqi Zhang⁴](#),
- [Xuemei Sun](#) ORCID: [orcid.org/0000-0002-2583-8593^{1,2,3}](https://orcid.org/0000-0002-2583-8593),
- [Bingjie Wang^{1,2,3}](#),
- [Peining Chen](#) ORCID: [orcid.org/0000-0003-0566-1660^{1,2,3}](https://orcid.org/0000-0003-0566-1660),
- [Yonggang Wang](#) ORCID: [orcid.org/0000-0002-2447-4679^{5,6,7}](https://orcid.org/0000-0002-2447-4679),
- [Yongyao Xia^{5,6,7}](#) &
- [Huisheng Peng](#) ORCID: [orcid.org/0000-0002-2142-2945^{1,2,3}](https://orcid.org/0000-0002-2142-2945)

Nature volume 597, pages 57–63 (2021)

- 4726 Accesses

- 50 Altmetric
- [Metrics details](#)

Subjects

- [Batteries](#)
- [Electrical and electronic engineering](#)

Abstract

Fibre lithium-ion batteries are attractive as flexible power solutions because they can be woven into textiles, offering a convenient way to power future wearable electronics^{1,2,3,4}. However, they are difficult to produce in lengths of more than a few centimetres, and longer fibres were thought to have higher internal resistances^{3,5} that compromised electrochemical performance^{6,7}. Here we show that the internal resistance of such fibres has a hyperbolic cotangent function relationship with fibre length, where it first decreases before levelling off as length increases. Systematic studies confirm that this unexpected result is true for different fibre batteries. We are able to produce metres of high-performing fibre lithium-ion batteries through an optimized scalable industrial process. Our mass-produced fibre batteries have an energy density of 85.69 watt hour per kilogram (typical values⁸ are less than 1 watt hour per kilogram), based on the total weight of a lithium cobalt oxide/graphite full battery, including packaging. Its capacity retention reaches 90.5% after 500 charge–discharge cycles and 93% at 1C rate (compared with 0.1C rate capacity), which is comparable to commercial batteries such as pouch cells. Over 80 per cent capacity can be maintained after bending the fibre for 100,000 cycles. We show that fibre lithium-ion batteries woven into safe and washable textiles by industrial rapier loom can wirelessly charge a cell phone or power a health management jacket integrated with fibre sensors and a textile display.

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: Measured and predicted internal resistances of FLIBs decrease with fibre length.

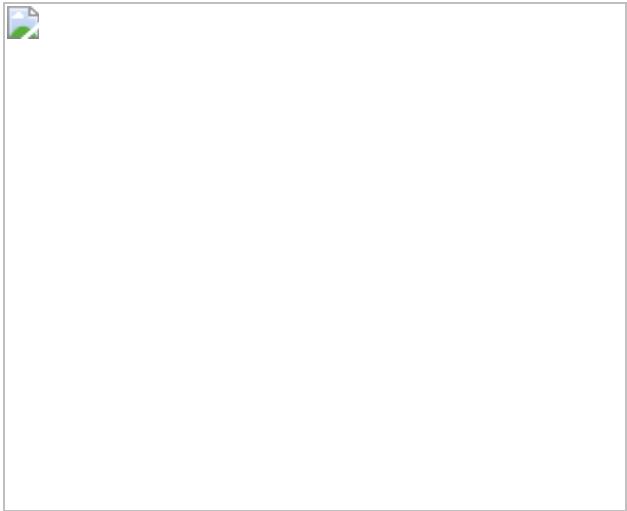


Fig. 2: Continuous fabrication and structural characterization of long FLIBs.

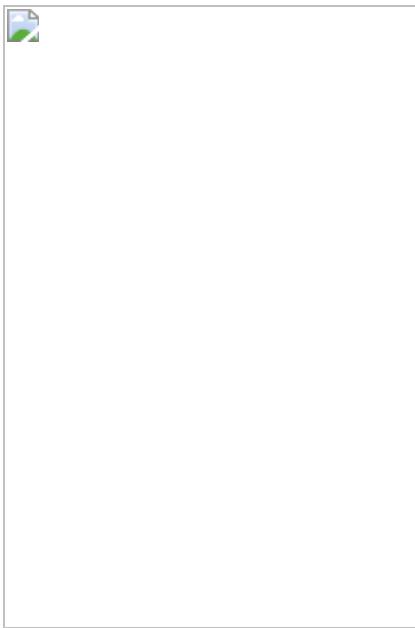


Fig. 3: Electrochemical properties of FLIBs.

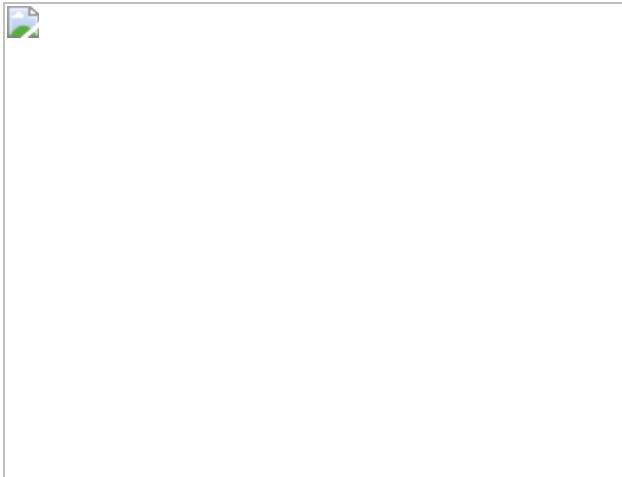
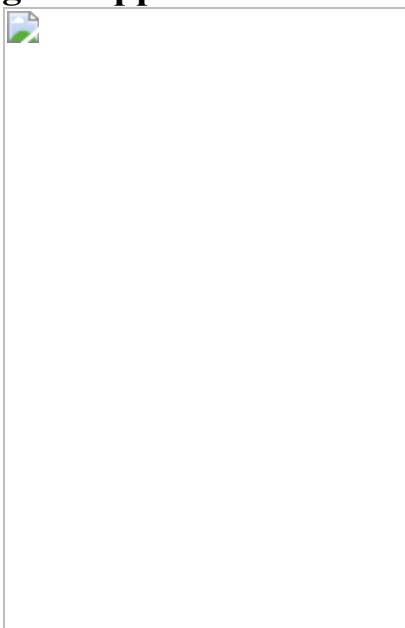


Fig. 4: Applications of FLIB textiles.



Data availability

The data that support the findings of this study are available from figshare at

https://figshare.com/articles/online_resource/Source_data_FLIBs/14775900. Source data are provided with this paper.

References

1. 1.

Mackanic, D. G., Kao, M. & Bao, Z. Enabling deformable and stretchable batteries. *Adv. Energy Mater.* **10**, 2001424 (2020).

2. 2.

Lee, J. et al. Recent advances in 1D stretchable electrodes and devices for textile and wearable electronics: materials, fabrications, and applications. *Adv. Mater.* **32**, 1902532 (2020).

3. 3.

Mo, F. et al. An overview of fiber-shaped batteries with a focus on multifunctionality, scalability, and technical difficulties. *Adv. Mater.* **32**, 1902151 (2020).

4. 4.

Sun, H. et al. Energy harvesting and storage in 1D devices. *Nat. Rev. Mater.* **2**, 17023 (2017).

5. 5.

Cheng, X. et al. Designing one-dimensional supercapacitors in a strip shape for high performance energy storage fabrics. *J. Mater. Chem. A* **3**, 19304–19309 (2015).

6. 6.

Kato, Y. et al. High-power all-solid-state batteries using sulfide superionic conductors. *Nat. Energy* **1**, 16030 (2016).

7. 7.

Chen, Y., Chang, K., Hu, C. & Cheng, T. Performance comparisons and resistance modeling for multi-segment electrode designs of power-oriented lithium-ion batteries. *Electrochim. Acta* **55**, 6433–6439 (2010).

8. 8.

Ren, J. et al. Twisting carbon nanotube fibers for both wire-shaped micro-supercapacitor and micro-battery. *Adv. Mater.* **25**, 1155–1159 (2013).

9. 9.

Tarascon, J. M. & Armand, M. Issues and challenges facing rechargeable lithium batteries. *Nature* **414**, 359–367 (2001).

10. 10.

Aubin, C. A. et al. Electrolytic vascular systems for energy-dense robots. *Nature* **571**, 51–57 (2019).

11. 11.

Zhang, X. et al. Two-dimensional MoS₂-enabled flexible rectenna for Wi-Fi-band wireless energy harvesting. *Nature* **566**, 368–372 (2019).

12. 12.

Han, C. G. et al. Giant thermopower of ionic gelatin near room temperature. *Science* **368**, 1091–1098 (2020).

13. 13.

Kwon, Y. H. et al. Cable-type flexible lithium ion battery based on hollow multi-helix electrodes. *Adv. Mater.* **24**, 5192–5197 (2012).

14. 14.

Wang, Y. et al. 3D-printed all-fiber Li-ion battery toward wearable energy storage. *Adv. Funct. Mater.* **27**, 1703140 (2017).

15. 15.

Ren, J. et al. Elastic and wearable wire-shaped lithium-ion battery with high electrochemical performance. *Angew. Chem. Int. Ed.* **53**, 7864–7869 (2014).

16. 16.

Wu, Z. et al. Ultrahigh-energy density lithium-ion cable battery based on the carbon-nanotube woven macrofilms. *Small* **14**, 1800414 (2018).

17. 17.

Wang, L. et al. A Li-air battery with ultralong cycle life in ambient air. *Adv. Mater.* **30**, 1704378 (2018).

18. 18.

Wang, K. et al. High-performance cable-type flexible rechargeable Zn battery based on MnO₂@CNT fiber microelectrode. *ACS Appl. Mater. Interfaces* **10**, 24573–24582 (2018).

19. 19.

Zhang, Y. et al. An ultraflexible silicon-oxygen battery fiber with high energy density. *Angew. Chem. Int. Ed.* **56**, 13741–13746 (2017).

20. 20.

Zhang, Y. et al. A fiber-shaped aqueous lithium ion battery with high power density. *J. Mater. Chem. A Mater. Energy Sustain.* **4**, 9002–9008 (2016).

21. 21.

Lee, J. et al. Reversible Mn²⁺/Mn⁴⁺ double redox in lithium-excess cathode materials. *Nature* **556**, 185–190 (2018).

22. 22.

Khan, Z. et al. Steady flow and heat transfer analysis of Phan-Thein-Tanner fluid in double-layer optical fiber coating analysis with slip conditions. *Sci. Rep.* **6**, 34593 (2016).

23. 23.

Mirri, F. et al. Lightweight, flexible, high-performance carbon nanotube cables made by scalable flow coating. *ACS Appl. Mater. Interfaces* **8**, 4903–4910 (2016).

24. 24.

Fan, H. et al. Continuously processed, long electrochromic fibers with multi-environmental stability. *ACS Appl. Mater. Interfaces* **12**, 28451–28460 (2020).

25. 25.

Quéré, D. Fluid coating on a fiber. *Annu. Rev. Fluid Mech.* **31**, 347–384 (1999).

26. 26.

de Ryck, A. & Quéré, D. Fluid coating from a polymer solution. *Langmuir* **14**, 1911–1914 (1998).

27. 27.

Li, G. X. et al. Stable metal battery anodes enabled by polyethylenimine sponge hosts by way of electrokinetic effects. *Nat. Energy* **3**, 1076–1083 (2018).

28. 28.

Niu, C. J. et al. Self-smoothing anode for achieving high-energy lithium metal batteries under realistic conditions. *Nat. Nanotechnol.* **14**, 594–601 (2019).

29. 29.

Wang, L. et al. Weaving sensing fibers into electrochemical fabric for real-time health monitoring. *Adv. Funct. Mater.* **28**, 1804456 (2018).

30. 30.

LeGrys, V. A. Sweat testing for the diagnosis of cystic fibrosis: practical considerations. *J. Pediatr.* **129**, 892–897 (1996).

31. 31.

Gao, W. et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature* **529**, 509–514 (2016).

32. 32.

Nyein, H. Y. et al. A Wearable electrochemical platform for noninvasive simultaneous monitoring of Ca^{2+} and pH. *ACS Nano* **10**, 7216–7224 (2016).

33. 33.

Kim, J., Campbell, A. S., de Avila, B. E. & Wang, J. Wearable biosensors for healthcare monitoring. *Nat. Biotechnol.* **37**, 389–406 (2019).

Acknowledgements

This work was supported by Ministry of Science and Technology of the People's Republic of China (2016YFA0203302), National Natural Science Foundation of China (21634003, 22075050, 21805044), Science and Technology Commission of Shanghai Municipality (20JC1414902, 18QA1400700, 19QA1400800) and Shanghai Municipal Education Commission (2017-01-07-00-07-E00062). We thank Ai Lin Chun of Science Storylab for critically reading and editing the manuscript.

Author information

Author notes

1. These authors contributed equally: Jiqing He, Chenhao Lu

Affiliations

1. State Key Laboratory of Molecular Engineering of Polymers, Fudan University, Shanghai, China

Jiqing He, Chenhao Lu, Haibo Jiang, Xiang Shi, Jingxia Wu, Liyuan Wang, Taiqiang Chen, Jiajia Wang, Ye Zhang, Han Yang, Xuemei Sun, Bingjie Wang, Peining Chen & Huisheng Peng

2. Department of Macromolecular Science, Fudan University, Shanghai, China

Jiqing He, Chenhao Lu, Haibo Jiang, Xiang Shi, Jingxia Wu, Liyuan Wang, Taiqiang Chen, Jiajia Wang, Ye Zhang, Han Yang, Xuemei Sun, Bingjie Wang, Peining Chen & Huisheng Peng

3. Laboratory of Advanced Materials, Fudan University, Shanghai, China

Jiqing He, Chenhao Lu, Haibo Jiang, Xiang Shi, Jingxia Wu, Liyuan Wang, Taiqiang Chen, Jiajia Wang, Ye Zhang, Han Yang, Xuemei Sun, Bingjie Wang, Peining Chen & Huisheng Peng

4. Institute of Future Lighting, Academy for Engineering and Technology, Fudan University, Shanghai, China

Fei Han & Guoqi Zhang

5. Department of Chemistry, Fudan University, Shanghai, China

Yonggang Wang & Yongyao Xia

6. Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Fudan University, Shanghai, China

Yonggang Wang & Yongyao Xia

7. iChEM (Collaborative Innovation Center of Chemistry for Energy Materials), Fudan University, Shanghai, China

Yonggang Wang & Yongyao Xia

Contributions

H.P. directed all aspects of the project. H.P. and P.C. conceived and designed the project. J.H. and C.L. performed the experiments on fabrication of FLIBs and integration systems. X.S. and J. Wu performed the fabrication of display textile. F.H., L.W., J. Wang and H.Y. performed the fabrication of fibre sensors. J.H., C.L. and H.J. analysed the data. Y.Z., T.C., G.Z., X.S., B.W., Y.W., Y.X. and all other authors discussed the data. J.H., C.L., P.C. and H.P. wrote the paper.

Corresponding authors

Correspondence to [Peining Chen](#) or [Huisheng Peng](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Kun Fu 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 Preliminary FLIBs.](#)

a, Schematic illustration of FLIB with a twisted structure. The typical active materials of LCO and graphite were coated on the aluminium and copper wires to form positive and negative fibre electrodes, respectively. The

negative fibre electrode was wrapped with commercial separator film to prevent short circuit. **b**, Photograph of a FLIB. Scale bar, 4 mm. **c**, Disassembled FLIB to show the components such as positive electrode, negative electrode, separator and encapsulation tube. Scale bar, 4 mm. **d**, Photograph of FLIBs with different lengths. **e**, Relationship between the internal resistance of FLIB and the battery length. Error bars: standard deviations of the results from three samples

[Source data](#).

Extended Data Fig. 2 Length dependence of internal resistance of FLIB under different conditions.

a–c, Length dependence of internal resistance for FLIBs using LCO/graphite (**a**), LiNi_{0.5}Co_{0.2}Mn_{0.3}O₂ (NCM523)/graphite (**b**), and LCO/SiO-graphite composites (SiC650) (**c**) as electrode materials. **d–f**, Length dependence of internal resistance of empty charged state (**d**), half charged state (**e**), and fully charged state (**f**) for LCO/graphite. **g–i**, Length dependence of internal resistance with increasing lithium salt concentrations of 0.01 M (**g**), 0.1 M (**h**) and 1 M (**i**) in electrolyte for LCO/graphite. Error bars: standard deviations of the results from three samples

[Source data](#).

Extended Data Fig. 3 Internal resistance of FLIB measured by EIS method.

a, Typical Nyquist plots of FLIBs with different lengths. The FLIBs show the same lithium-ion concentration of 1 M in electrolyte and charged state of 50%. **b**, Z values along the fibre length of an LCO/graphite FLIB. Here, L is the investigated length on a FLIB with a length of L_0 . Z was obtained by measuring the polarization resistance of FLIB through EIS analysis, that is, it was calculated as a sum of interface resistance and charge transfer resistance from the equivalent circuit simulation curve of the Nyquist plot. Error bars: standard deviations of the results from three samples. **c**, Internal

resistances of FLIBs obtained from EIS correlate linearly ($y = 1.06047x$; $R^2 = 0.99663$) with predicted values. Error bars: standard deviations of the results from three samples. **d–f**, Internal resistance of LCO/graphite (**d**), NCM523/graphite (**e**) and LCO/SiC650 (**f**) FLIBs as a function of length. **g–i**, Internal resistance of LCO/graphite FLIB at empty charged state (**g**), half charged state (**h**) and fully charged state (**i**). **j–l**, Internal resistance of LCO/graphite FLIB with increasing lithium salt concentrations of 0.01 M (**j**), 0.1 M (**k**) and 1 M (**l**) in electrolyte. Error bars: standard deviations of the results from three samples

[Source data](#).

Extended Data Fig. 4 Continuous coating of fibre electrodes.

a, Binder ratio dependence of deformation parameter for the beads on the fibre electrode. With increasing binder contents, the beads disappeared, and the electrode surface turned smooth. Deformation parameter $D = (L - B)/(L+B)$, where L is bead size along the fibre and B is bead size in the radial direction of the fibre. Scale bar, 500 μm . **b**, Photographs of fluid coating process. (i) Photograph of fibre fluid coating machine. Scale bar, 30 cm. (ii–iv) Photographs of detailed coating processes. Scale bar, 5 cm. **c**, **d**, SEM images show that bending of positive (**c**) and negative (**d**) fibre electrodes with a curvature radius of 1.5 mm did not cause any peeling or cracks in the active layer. Scale bars, from left to right, 2 mm (i), 500 μm (ii) and 100 μm (iii). **e**, **f**, SEM images of positive and negative fibre electrode after bending for 100,000 cycles with a curvature radius of 1 cm. Scale bar, 500 μm . **g**, LCO loading weight and corresponding discharge capacity retention (1 C rate) for five different electrodes. Capacity retention remained stable at first and then decreased with increasing LCO loading weight. Error bars: standard deviations of the results from three samples. **h**, **i**, Charge–discharge profiles of LCO/Li (**h**) and graphite/Li (**i**) of coin cells

[Source data](#).

Extended Data Fig. 5 Assembly of fibre electrodes.

a–c, Separator wrapping angle of 27° (**a**), 22° (**b**) and 16° (**c**). Scale bars, 10 cm. **d–f**, SEM images of wrapped negative fibres with wrapping angles of 27° (**d**), 22° (**e**) and 16° (**f**). Scale bars, 500 µm. When the wrapping angle was 16°, the negative fibre electrode was fully covered by the separator strip without excessive overlap. **g–j**, Photographs of twisted fibre electrodes with pitches of 4 mm (**g**), 3 mm (**h**), 2 mm (**i**) and 1.5 mm (**j**). Scale bar, 4 mm. **k–n**, SEM images of twisted fibre electrodes with pitches of 4 mm (**k**), 3 mm (**l**), 2 mm (**m**) and 1.5 mm (**n**). **o**, Internal resistances of 20-cm-long FLIBs with different twisting pitches. Error bars represent the standard deviations of the results from three samples. **p**, Capacity retention of FLIBs with pitches from 4 mm to 1.5 mm at different applied discharge rate. FLIBs with a large pitch cause high internal resistance, but too small a pitch would result in peeling of active materials. **q**, Photograph of twisted fibre electrodes with a pitch of 2 mm. Scale bar, 2 mm. **r–u**, Statistical distributions for twisted fibre electrodes with different pitches of 4 mm (**r**), 3 mm (**s**), 2 mm (**t**) and 1.5 mm (**u**)

Source data.

Extended Data Fig. 6 Encapsulation of FLIB.

a, Illustration of extruder. The right top and bottom photographs show twisted fibre electrodes being inserted into extrusion die and encapsulating electrodes through extruding polymer composite tube. Scale bar, 2 cm. **b**, Cross-sectional SEM image of fibre electrodes with encapsulation tube, which had an outer diameter of 1.5 mm with tube thickness of 0.25 mm. Scale bar, 500 µm. **c**, Schematic showing the injection pump and the tip-forming machine. **d**, Schematic showing the set-up of the tip-forming machine used to encapsulate the end of the FLIBs. One end of the FLIB is inserted into the sealing hole (i, ii) and heated at 300 °C for 5 s (iii). Upon cooling, a smooth and solid end is formed (iv). **e**, Photograph of the end of an encapsulated tube. Electrolyte is in red. Scale bar, 1 cm. **f**, No obvious electrolyte leakage from the end of the encapsulated tube was observed when submerged in water. Scale bar, 1 cm. **g**, Tensile strength of the encapsulation tube remained largely unchanged after being immersed in the electrolyte for different durations. σ_0 and σ represent tensile strengths before and after immersion, respectively. Error bars are standard deviations of the

results from five samples. **h**, Photograph of the standard washing machine and washing container for the washing test. Scale bars, from left to right, 20 cm and 5 cm. **i**, Photograph of drying equipment. Scale bar, 30 cm. **j**, Tensile strength of the encapsulation tube remained unchanged after several washing and drying cycles. σ_0 and σ represent the tensile strengths before and after washing, respectively. Error bars are standard deviations of the results from five samples

[Source data](#).

Extended Data Fig. 7 Test report of FLIB.

a, b, Scanned copies of the independent confirmation report (**a**) and appendix (**b**) showing the capacity and energy density of FLIB. Discharge capacity is required by the testing standard authorized by China National Accreditation Service for Conformity Assessment (CNAS) and China Inspection Body and Laboratory Mandatory Approval (CMA), while discharge energy is not required. The remark “The discharge energy test is beyond the independent confirmation range of CNAS and CMA” in the independent confirmation report simply indicates that discharge energy test is not a requirement within the testing standard.

Extended Data Fig. 8 Electrochemical stability and durability of FLIBs.

a, Capacity retention of FLIBs from -20°C to 60°C . Error bars represent the standard deviations of the results from three samples. **b**, Stress–strain curve shows an LCO/graphite FLIB has a tensile strength of around 89 MPa and elongation of around 13%. **c**, Photographs of a FLIB being dynamically bent at a frequency of 2.5 Hz while its open-circuit voltage was traced during the bending cycle. **d, e**, Charge and discharge profiles of the FLIB under dynamic bending (at 2.5 Hz) are stable with no obvious fluctuations. **f**, Capacity retention of gel-electrolyte-based FLIB reaching $>90\%$ after 100,000 bending cycles. Error bars: standard deviations of the results from three samples. Insets: charge–discharge profiles at the 1st, 100th and 100,000th bending cycle

[Source data](#).

Extended Data Fig. 9 Fabrication of FLIB textile, and FLIB textile charging a tablet computer under harsh conditions.

a–d, Textile made from FLIBs by a commercial rapier loom. **a, b**, Fabrication process using a rapier loom (inset, rolling-up of a textile). Scale bar, 10 cm. **c**, Photograph of a 5-m-long, 0.3-m-wide FLIB textile. **d**, Photograph of FLIB textile with different styles. Scale bar, 2 cm. **e**, FLIB textile charging a tablet as normal, even when folded. **i**, FLIB textile charging tablet; **ii**, charging tablet after two folds; **iii**, charging tablet after three folds. The size of the textile was 0.3 m long × 0.5 m wide. **f**, FLIB textile charging tablet under car crushing. The car mass was about 1,300 kg. **g**, FLIB textile being washed by a washing machine. Scale bar, 10 cm. **h**, FLIB textile charging a tablet after washed and dried. **i**, FLIB textile charging tablet when punctured by a blade. **j, k**, The temperature of the zone monitored by the infrared imager remains almost unchanged before (**j**) and after puncture (**k**). **l**, Photograph of FLIB textile wirelessly charging a cell phone at 40 °C. FLIB textile was heated with a hot stage. **m**, Infrared thermal images of FLIB textile during wireless charging of cell phone.

Extended Data Fig. 10 Fabrication of a health management system integrated with fibre sweat sensors and an electroluminescent textile display.

a, b, Sweat fibre sensors integrated with Na^+ , Ca^{2+} working fibre electrodes and Ag/AgCl reference fibre electrode. Scale bar, 1 cm. **c, d**, Calibration plot for Na^+ (**c**) and Ca^{2+} (**d**). The ion concentration was converted from the potential signal according to the calibration plot. **e, f**, Photographs of the sewing machine (**e**) and electroluminescent yarn (**f**) used to fabricate electroluminescent textile display. Scale bar, 5 cm

[Source data](#).

Extended Data Table 1 Key performances of FLIB and typical commercial batteries

Supplementary information

Supplementary Information

This file contains Supplementary Figures 1 – 3 and Supplementary Tables 1 – 6.

Supplementary Video 1

Production process of FLIBs.

Supplementary Video 2

FLIB textile wirelessly charges a smart phone.

Supplementary Video 3

FLIB textile stably charges a pad under folding.

Supplementary Video 4

FLIB textile normally charges a pad after washing.

Supplementary Video 5

FLIB textile normally charges a pad when punctured by blade.

Supplementary Video 6

A jacket woven with FLIBs wirelessly charges a smart phone.

Source data

Source Data Fig. 1

[**Source Data Fig. 2**](#)

[**Source Data Fig. 3**](#)

[**Source Data Fig. 4**](#)

[**Source Data Extended Data Fig. 1**](#)

[**Source Data Extended Data Fig. 2**](#)

[**Source Data Extended Data Fig. 3**](#)

[**Source Data Extended Data Fig. 4**](#)

[**Source Data Extended Data Fig. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

[**Source Data Extended Data Fig. 8**](#)

[**Source Data Extended Data Fig. 10**](#)

Rights and permissions

[**Reprints and Permissions**](#)

About this article

Cite this article

He, J., Lu, C., Jiang, H. *et al.* Scalable production of high-performing woven lithium-ion fibre batteries. *Nature* **597**, 57–63 (2021).
<https://doi.org/10.1038/s41586-021-03772-0>

- Received: 23 February 2021
 - Accepted: 25 June 2021
 - Published: 01 September 2021
 - Issue Date: 02 September 2021
 - DOI: <https://doi.org/10.1038/s41586-021-03772-0>
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03772-0>

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

- Article
- [Published: 19 July 2021](#)

Cleaving arene rings for acyclic alkenylnitrile synthesis

- [Xu Qiu](#) [ORCID: orcid.org/0000-0002-9559-3624](#)^{1 na1},
- [Yueqian Sang](#)^{3 na1},
- [Hao Wu](#)^{1,4 na1},
- [Xiao-Song Xue](#) [ORCID: orcid.org/0000-0003-4541-8702](#)^{2,3},
- [Zixi Yan](#)¹,
- [Yachong Wang](#)¹,
- [Zengrui Cheng](#)¹,
- [Xiaoyang Wang](#)¹,
- [Hui Tan](#)¹,
- [Song Song](#)¹,
- [Guisheng Zhang](#)⁴,
- [Xiaohui Zhang](#)¹,
- [K. N. Houk](#) [ORCID: orcid.org/0000-0002-8387-5261](#)² &
- [Ning Jiao](#) [ORCID: orcid.org/0000-0003-0290-9034](#)^{1,5}

[Nature](#) volume 597, pages 64–69 (2021)

- 12k Accesses
- 29 Altmetric
- [Metrics details](#)

Subjects

- [Homogeneous catalysis](#)

- [Synthetic chemistry methodology](#)
- [Chemical synthesis](#)
- [Theoretical chemistry](#)

Abstract

Synthetic chemistry is built around the formation of carbon–carbon bonds. However, the development of methods for selective carbon–carbon bond cleavage is a largely unmet challenge^{1,2,3,4,5,6}. Such methods will have promising applications in synthesis, coal liquefaction, petroleum cracking, polymer degradation and biomass conversion. For example, aromatic rings are ubiquitous skeletal features in inert chemical feedstocks, but are inert to many reaction conditions owing to their aromaticity and low polarity. Over the past century, only a few methods under harsh conditions have achieved direct arene-ring modifications involving the cleavage of inert aromatic carbon–carbon bonds^{7,8}, and arene-ring-cleavage reactions using stoichiometric transition-metal complexes or enzymes in bacteria are still limited^{9,10,11}. Here we report a copper-catalysed selective arene-ring-opening reaction strategy. Our aerobic oxidative copper catalyst converts anilines, arylboronic acids, aryl azides, aryl halides, aryl triflates, aryl trimethylsiloxanes, aryl hydroxamic acids and aryl diazonium salts into alkenyl nitriles through selective carbon–carbon bond cleavage of arene rings. This chemistry was applied to the modification of polycyclic aromatics and the preparation of industrially important hexamethylenediamine and adipic acid derivatives. Several examples of the late-stage modification of complex molecules and fused ring compounds further support the potential broad utility of this methodology.

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: Cleaving arene rings.

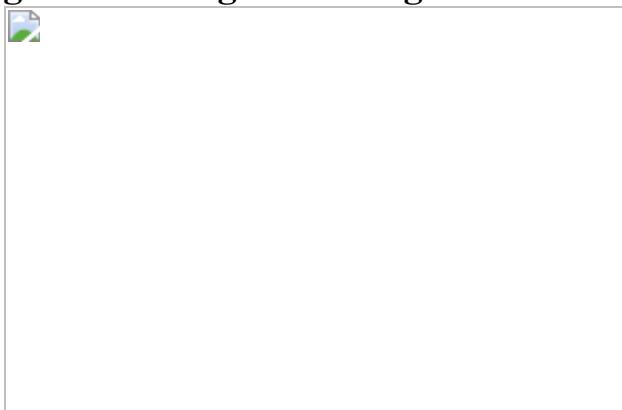


Fig. 2: Scope of polycyclic aromatic compounds and transformations to *ortho*(*cis*-cyanovinyl) aryl nitriles.

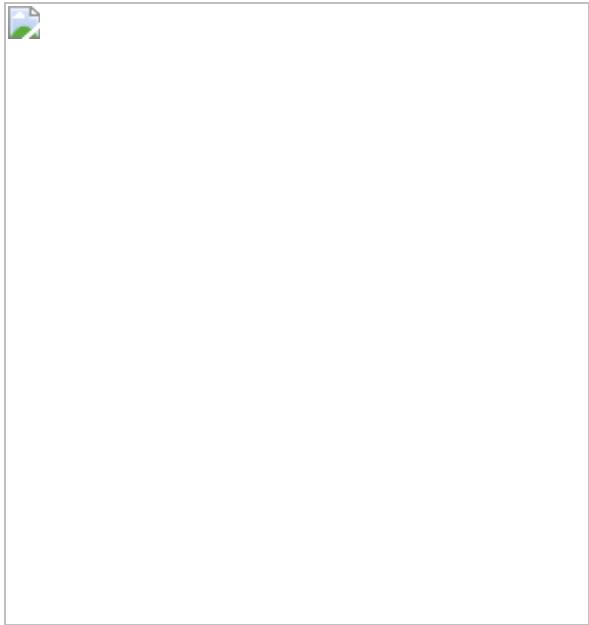


Fig. 3: The cleavage of anilines and phenylboronic acids and downstream transformations.

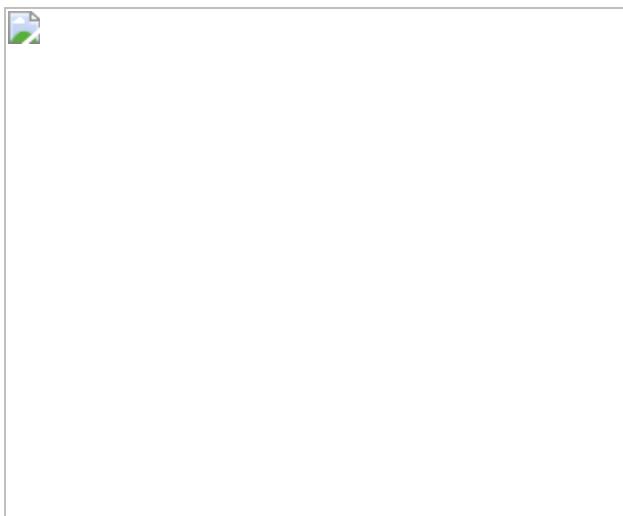
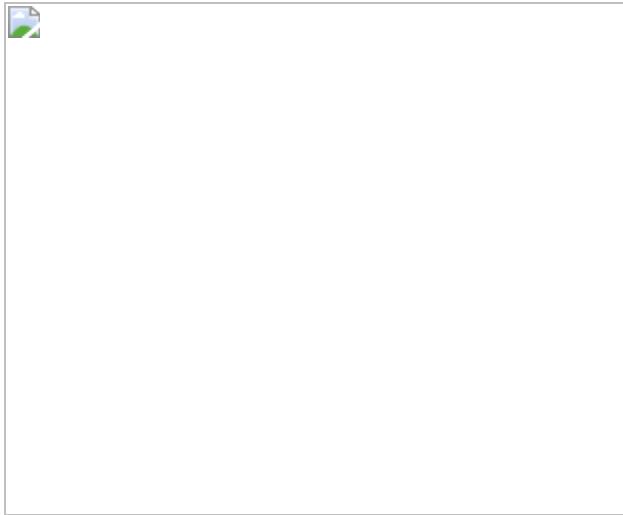


Fig. 4: Mechanism studies.



Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

References

1. 1.

National Research Council (US) *Health and Medicine: Challenges for the Chemical Sciences in the 21st Century* (National Academies Press, 2004).

2. 2.

Jones, W. D. The fall of the C–C bond. *Nature* **364**, 676–677 (1993).

3. 3.

Zhu, J., Wang, J. & Dong, G. Catalytic activation of unstrained C(aryl)–C(aryl) bonds in 2,2'-biphenols. *Nat. Chem.* **11**, 45–51 (2019).

4. 4.

Guengerich, F. P. & Yoshimoto, F. K. Formation and cleavage of C–C Bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655 (2018).

5. 5.

Jakoobi, M. & Sergeev, A. G. Transition-metal-mediated cleavage of C–C bonds in aromatic rings. *Chem. Asian J.* **14**, 2181–2192 (2019).

6. 6.

Murakami, M. & Chatani, N. *Cleavage of Carbon–Carbon Single Bonds by Transition Metals* (Wiley, 2015).

7. 7.

Mortier, J. *Arene Chemistry: Reaction Mechanisms and Methods for Aromatic Compounds* (Wiley, 2015).

8. 8.

Benitez, V. M., Grau, J. M., Yori, J. C., Pieck, C. L. & Vera, C. R. Hydroisomerization of benzene-containing paraffinic feedstocks over Pt/WO₃–ZrO₂ catalysts. *Energy Fuels* **20**, 1791–1798 (2006).

9. 9.

Sattler, A. & Parkin, G. Cleaving carbon–carbon bonds by inserting tungsten into unstrained aromatic rings. *Nature* **463**, 523–526 (2010).

10. 10.

Hu, S., Shima, T. & Hou, Z. Carbon–carbon bond cleavage and rearrangement of benzene by a trinuclear titanium hydride. *Nature* **512**, 413–415 (2014).

11. 11.

Bugg, T. D. H. & Winfield, C. J. Enzymatic cleavage of aromatic rings: mechanistic aspects of the catechol dioxygenases and later enzymes of bacterial oxidative cleavage pathways. *Nat. Prod. Rep.* **15**, 513–530 (1998).

12. 12.

Wilson, J. Celebrating Michael Faraday’s discovery of benzene. *Ambix* **59**, 241–265 (2013).

13. 13.

Labinger, J. A. & Bercaw, J. E. Understanding and exploiting C–H bond activation. *Nature* **417**, 507–514 (2002).

14. 14.

Gensch, T., Hopkinson, M. N., Glorius, F. & Wencel-Delord, J. Mild metal-catalyzed C–H activation: examples and concepts. *Chem. Soc. Rev.* **45**, 2900–2936 (2016).

15. 15.

Wang, Y., Li, J. & Liu, A. Oxygen activation by mononuclear nonheme iron dioxygenases involved in the degradation of aromatics. *J. Biol. Inorg. Chem.* **22**, 395–405 (2017).

16. 16.

Siddiqi, Z., Wertjes, W. C. & Sarlah, D. Chemical equivalent of arene monooxygenases: dearomatic synthesis of arene oxides and oxepines. *J. Am. Chem. Soc.* **142**, 10125–10131 (2020).

17. 17.

Kong, R. Y. & Crimmin, M. R. Chemoselective C–C σ -bond activation of the most stable ring in biphenylene. *Angew. Chem. Int. Ed.* **60**, 2619–2623 (2021).

18. 18.

Nakagawa, K. & Onoue, H. Oxidation of *o*-phenylenediamines with lead tetra-acetate. *Chem. Commun. (London)* **396a** (1965).

19. 19.

Nakagawa, K. & Onoue, H. Oxidation with nickel peroxide. V. The formation of *cis,cis*-1,4-dicyano-1,3-butadienes in the oxidation of *o*-phenylenediamines. *Tetrahedr. Lett.* **6**, 1433–1436 (1965).

20. 20.

Kajimoto, T., Takahashi, H. & Tsuji, J. Copper-catalyzed oxidation of *o*-phenylenediamines to *cis,cis*-mucononitriles. *J. Org. Chem.* **41**, 1389–1393 (1976).

21. 21.

Buchner, E. & Curtius, T. Synthese von Ketonsäureäthern aus Aldehyden und Diazoessigäther. *Ber. Dtsch. Chem. Ges.* **18**, 2371–2377 (1885).

22. 22.

Chapman, O. L. & Leroux, J. P. 1-Aza-1,2,4,6-cycloheptatetraene. *J. Am. Chem. Soc.* **100**, 282–285 (1978).

23. 23.

Satake, K., Mizushima, H., Kimura, M. & Morosawa, S. The reactions of nitrene for the conjugated π -systems. *Heterocycles* **23**, 195 (1985).

24. 24.

Liu, L. L. et al. A transient vinylphosphinidene via a phosphirene-phosphinidene rearrangement. *J. Am. Chem. Soc.* **140**, 147–150 (2018).

25. 25.

Hall, J. H. Dinitrenes from *o*-diazides. synthesis of 1,4-dicyano-1,3-butadienes. *J. Am. Chem. Soc.* **87**, 1147–1148 (1965).

26. 26.

Campbell, C. D. & Rees, C. W. Oxidation of 1- and 2-aminobenzotriazole. *Chem. Commun. (London)* 192–193 (1965).

27. 27.

Nicolaides, A. et al. Of *ortho*-conjugatively linked reactive intermediates: the cases of *ortho*-phenylene-(bis)nitrene, -carbenonitrene, and -(bis)carbene. *J. Am. Chem. Soc.* **121**, 10563–10572 (1999).

28. 28.

Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).

29. 29.

Chen, Y., Kamlet, A. S., Steinman, J. B. & Liu, D. R. A biomolecule-compatible visible-light-induced azide reduction from a DNA-encoded reaction-discovery system. *Nat. Chem.* **3**, 146–153 (2011).

30. 30.

Prescher, J. A., Dube, D. H. & Bertozzi, C. R. Chemical remodelling of cell surfaces in living animals. *Nature* **430**, 873–877 (2004).

31. 31.

Haag, S. M. et al. Targeting STING with covalent small-molecule inhibitors. *Nature* **559**, 269–273 (2018).

32. 32.

Meng, G. et al. Modular click chemistry libraries for functional screens using a diazotizing reagent. *Nature* **574**, 86–89 (2019).

33. 33.

Sharma, A. & Hartwig, J. F. Metal-catalysed azidation of tertiary C–H bonds suitable for late-stage functionalization. *Nature* **517**, 600–604 (2015).

34. 34.

Zhdankin, V. V. et al. Preparation, X-ray crystal structure, and chemistry of stable azidoiodinanes derivatives of benziodoxole. *J. Am. Chem. Soc.* **118**, 5192–5197 (1996).

35. 35.

Huang, X., Bergsten, T. M. & Groves, J. T. Manganese-catalyzed late-stage aliphatic C–H azidation. *J. Am. Chem. Soc.* **137**, 5300–5303 (2015).

36. 36.

Schmidt, K. F. Über die Einwirkung von NH auf organische Verbindungen. *Angew. Chem.* **36**, 511 (1923).

37. 37.

Schmidt, K. F. Über den Imin-Rest. *Ber. Dtsch. Chem. Ges.* **57B**, 704–706 (1924).

38. 38.

Liu, J. et al. From alkylarenes to anilines via site-directed carbon–carbon amination. *Nat. Chem.* **11**, 71–77 (2018).

39. 39.

Fu, N., Sauer, G. S., Saha, A., Loo, A. & Lin, S. Metal-catalyzed electrochemical diazidation of alkenes. *Science* **357**, 575–579 (2017).

40. 40.

Scriven, E. F. V. & Turnbull, K. Azides: their preparation and synthetic uses. *Chem. Rev.* **88**, 297–368 (1988).

41. 41.

Lombardi, F. et al. Quantum units from the topological engineering of molecular graphenoids. *Science* **366**, 1107–1110 (2019).

42. 42.

Kolmer, M. et al. Fluorine-programmed nanozipping to tailored nanographenes on rutile TiO₂ surfaces. *Science* **363**, 57–60 (2019).

43. 43.

Yano, Y. et al. Living annulative pi-extension polymerization for graphene nanoribbon synthesis. *Nature* **571**, 387–392 (2019).

44. 44.

Hawker, C. J. & Wooley, K. L. The convergence of synthetic organic and polymer chemistries. *Science* **309**, 1200–1205 (2005).

45. 45.

Oh, J. Y. et al. Intrinsically stretchable and healable semiconducting polymer for organic transistors. *Nature* **539**, 411–415 (2016).

46. 46.

Bakhoda, A. G., Jiang, Q., Bertke, J. A., Cundari, T. R. & Warren, T. H. Elusive terminal copper arylnitrene intermediates. *Angew. Chem. Int. Ed.* **56**, 6426–6430 (2017).

47. 47.

Carsch, K. M. et al. Synthesis of a copper-supported triplet nitrene complex pertinent to copper-catalyzed amination. *Science* **365**, 1138–1143 (2019).

48. 48.

Allen, S. E., Walvoord, R. R., Padilla-Salinas, R. & Kozlowski, M. C. Aerobic copper-catalyzed organic reactions. *Chem. Rev.* **113**, 6234–6458 (2013).

49. 49.

McCann, S. D. & Stahl, S. S. Copper-catalyzed aerobic oxidations of organic molecules: pathways for two-electron oxidation with a four-electron oxidant and a one-electron redox-active catalyst. *Acc. Chem. Res.* **48**, 1756–1766 (2015).

Acknowledgements

We acknowledge the NSFC (grant numbers 21632001, 21772002, 81821004, 21933004), the National Key Research and Development Project (grant number 2019YFC1708902), and the US National Science Foundation (CHE-1764328) for financial support of this research.

Author information

Author notes

1. These authors contributed equally: Xu Qiu, Yueqian Sang, Hao Wu

Affiliations

1. State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, China

Xu Qiu, Hao Wu, Zixi Yan, Yachong Wang, Zengrui Cheng, Xiaoyang Wang, Hui Tan, Song Song, Xiaohui Zhang & Ning Jiao

2. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA

Xiao-Song Xue & K. N. Houk

3. State Key Laboratory of Elemento-organic Chemistry, College of Chemistry, Nankai University, Tianjin, China

Yueqian Sang & Xiao-Song Xue

4. Key Laboratory of Green Chemical Media and Reactions, Ministry of Education, School of Chemistry and Chemical Engineering, Henan Normal University, Xinxiang, China

Hao Wu & Guisheng Zhang

5. State Key Laboratory of Organometallic Chemistry, Chinese Academy of Sciences, Shanghai, China

Ning Jiao

Contributions

N.J. conceived the project and directed the research. K.N.H. and X.-S.X. supervised the mechanistic study. X.Q., Y.S., X.-S.X., K.N.H. and N.J. wrote the paper. X.Q., H.W., Z.Y., Y.W., Z.C. and X.W. performed the experiments. Y.S. performed the DFT calculations. H.T, S.S., G.Z. and X.Z. discussed the results.

Corresponding authors

Correspondence to [K. N. Houk](#) or [Ning Jiao](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Adrian Mulholland 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 Downstream transformations and mechanism studies.

a, Downstream transformations of alkenyl nitriles. **b**, The excluded intermediates. **c**, HOMO(α) and HOMO-1(β) of the triplet copper bis-nitrene intermediate. *See [Supplementary Information](#) for experimental details.

Supplementary information

Supplementary Information

This file contains Supplementary Information (see Table of Contents in PDF for full description).

Supplementary Data

This file contains the crystal structure of **48** in Fig 3.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Qiu, X., Sang, Y., Wu, H. *et al.* Cleaving arene rings for acyclic alkenylnitrile synthesis. *Nature* **597**, 64–69 (2021).
<https://doi.org/10.1038/s41586-021-03801-y>

- Received: 17 February 2021
- Accepted: 06 July 2021
- Published: 19 July 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03801-y>

[Benzene rings broken for chemical synthesis](#)

- Mark R. Crimmin

News & Views 01 Sept 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03801-y>

- Article
- [Published: 01 September 2021](#)

Enantioselective synthesis of ammonium cations

- [Mark P. Walsh ORCID: orcid.org/0000-0003-3185-2460¹](#),
- [Joseph M. Phelps ORCID: orcid.org/0000-0002-6706-8742¹](#),
- [Marc E. Lennon ORCID: orcid.org/0000-0003-1947-4348¹](#),
- [Dmitry S. Yufit¹](#) &
- [Matthew O. Kitching ORCID: orcid.org/0000-0001-8276-583X¹](#)

Nature volume 597, pages 70–76 (2021)

- 4875 Accesses
- 83 Altmetric
- [Metrics details](#)

Subjects

- [Asymmetric synthesis](#)
- [Synthetic chemistry methodology](#)
- [Stereochemistry](#)
- [Supramolecular chemistry](#)

Abstract

Control of molecular chirality is a fundamental challenge in organic synthesis. Whereas methods to construct carbon stereocentres enantioselectively are well established, routes to synthesize enriched

heteroatomic stereocentres have garnered less attention^{1,2,3,4,5}. Of those atoms commonly present in organic molecules, nitrogen is the most difficult to control stereochemically. Although a limited number of resolution processes have been demonstrated^{6,7,8}, no general methodology exists to enantioselectively prepare a nitrogen stereocentre. Here we show that control of the chirality of ammonium cations is easily achieved through a supramolecular recognition process. By combining enantioselective ammonium recognition mediated by 1,1'-bi-2-naphthol scaffolds with conditions that allow the nitrogen stereocentre to racemize, chiral ammonium cations can be produced in excellent yields and selectivities. Mechanistic investigations demonstrate that, through a combination of solution and solid-phase recognition, a thermodynamically driven adductive crystallization process is responsible for the observed selectivity. Distinct from processes based on dynamic and kinetic resolution, which are under kinetic control, this allows for increased selectivity over time by a self-corrective process. The importance of nitrogen stereocentres can be revealed through a stereoselective supramolecular recognition, which is not possible with naturally occurring pseudoenantiomeric *Cinchona* alkaloids. With practical access to the enantiomeric forms of ammonium cations, this previously ignored stereocentre is now available to be explored.

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: Nitrogen stereocentres.

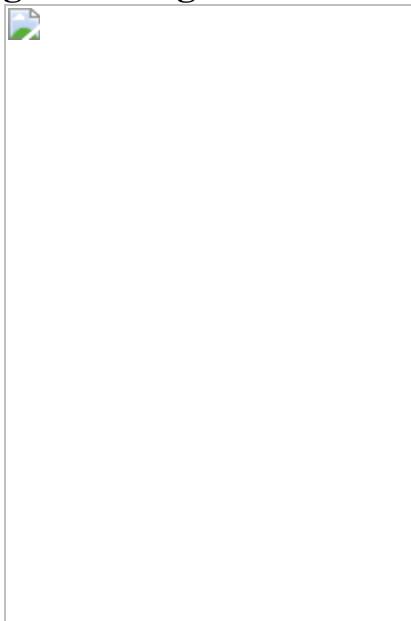


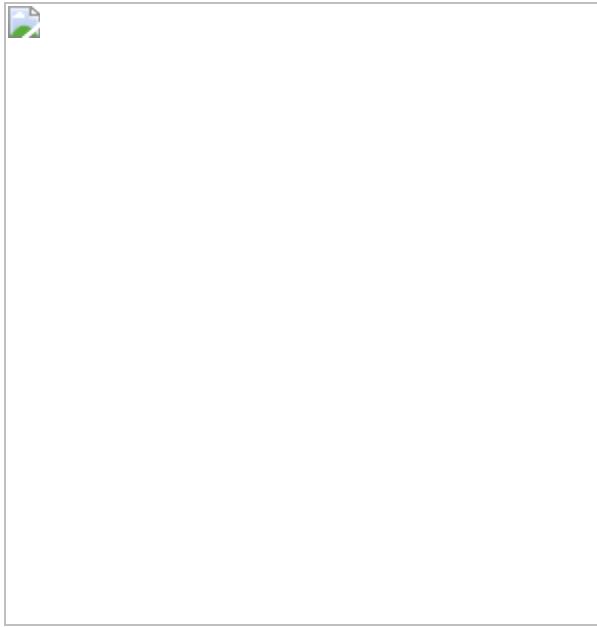
Fig. 2: General enantioselective ammonium recognition.



Fig. 3: Dynamic behaviour of ammonium cations.



Fig. 4: Enantioselective synthesis of ammonium cations.



Data Availability

Full crystallographic details in CIF format have been deposited in the Cambridge Crystallographic Data Centre database (deposition numbers: CCDC-1987042–1987058; 1987061–1987068; 1987165–1987180; 2047299–2047303). All other data are available from the corresponding author upon request.

References

1. 1.

Knouse, K. W. et al. Unlocking P(V): reagents for chiral phosphorothioate synthesis. *Science* **361**, 1234–1238 (2018).

2. 2.

Hayakawa, Y., Hyodo, M., Kimura, K. & Kataoka, M. The first asymmetric synthesis of trialkyl phosphates on the basis of dynamic kinetic resolution in the phosphite method using a chiral source in a catalytic manner. *Chem. Commun.* 1704–1705 (2003).

3. 3.

Bergin, E. et al. Synthesis of P-stereogenic phosphorus compounds. Asymmetric oxidation of phosphines under Appel conditions. *J. Am. Chem. Soc.* **129**, 9566–9567 (2007).

4. 4.

Pitchen, P., Duñach, E., Deshmukh, M. N. & Kagan, H. B. An efficient asymmetric oxidation of sulfides to sulfoxides. *J. Am. Chem. Soc.* **106**, 8188–8193 (1984).

5. 5.

Liu, G., Cogan, D. A. & Ellman, J. A. Catalytic asymmetric synthesis of *tert*-butanesulfinamide. Application to the asymmetric synthesis of amines. *J. Am. Chem. Soc.* **119**, 9913–9914 (1997).

6. 6.

Wedekind, E. Zur charaktersitik stereoisomerer Ammoniumsaize. *J. Chem. Soc.* **32**, 3561–3569 (1899).

7. 7.

Wedekind, E. & Wedekind, O. Über die Aktivierung einer cyclischen asymmetrischen Ammoniumbase. *Chem. Ber.* **40**, 4450–4456 (1907).

8. 8.

Fröhlich, E. & Wedekind, E. Über asymmetrische ammoniumsaize des *p*-anisidins. *Chem. Ber.* **40**, 1009–1013 (1907).

9. 9.

Lehn, J.-M. Nitrogen inversion. *Fortschr. Chem. Forsch.* **15**, 311–377 (1970).

10. 10.

Dolling, U.-H., Davis, P. & Grabowski, E. J. J. Efficient catalytic asymmetric alkylations. 1. Enantioselective synthesis of (+)-indacrinone via chiral phase-transfer catalysis. *J. Am. Chem. Soc.* **106**, 446–447 (1984).

11. 11.

Brown, D. R., Lygo, R., McKenna, J., McKenna, J. M. & Hutley, B. G. The preferred steric course of quaternisation of 1-alkylpiperidines. *J. Chem. Soc.* **1967** 1184–1194 (1967).

12. 12.

Brois, S. J. Aziridines. XI. Nitrogen inversion in N-haloaziridines. *J. Am. Chem. Soc.* **90**, 506–508 (1968).

13. 13.

Montanari, F., Moretti, I. & Torre, G. Asymmetric introduction at trivalent nitrogen. Optically active 2-methyl-3,3-diphenyloxaziridine, a compound with molecular asymmetry due solely to the nitrogen atom. *Chem. Commun.* **1968**, 1694–1695 (1968).

14. 14.

Mannschreck, A. & Seitz, W. Separation of invertomers (diastereoisomers) of diaziridines. Slow inversion at trivalent nitrogen atoms. *Angew. Chem. Int. Edn Engl.* **8**, 212–213 (1969).

15. 15.

Prelog, V. & Wieland, P. Über die Spaltung der Tröger'schen Base in optische Antipoden, ein Beitrag zur Stereochemie des dreiwertigen Stickstoffs. *Helv. Chim. Acta* **27**, 1127–1134 (1944).

16. 16.

Wilen, S. H., Qi, J. Z. & Williard, P. G. Resolution, asymmetric transformation, and configuration of Troeger's base. Application of

Troeger's base as a chiral solvating agent. *J. Org. Chem.* **56**, 485–487 (1991).

17. 17.

Pope, W. J. & Peachey, S. J. Asymmetric optically active nitrogen compounds. Dextro- and laevo-benzylphenylallylmethylammonium iodides and bromides. *J. Chem. Soc.* **75**, 1127–1131 (1899).

18. 18.

Havinga, E. Spontaneous formation of optically active substances. *Biochim. Biophys. Acta* **13**, 171–174 (1954).

19. 19.

Kostyanovsky, R. G., Lyssenko, K. A., Krutiusa, O. N. & Kostyanovsky, V. R. Isomorphism of chiral ammonium salts Ph(All)N⁺Et(Me)X⁻CHCl₃. *Mendeleev Commun.* **19**, 19–20 (2009).

20. 20.

Torbeev, V. Y., Lyssenko, K. A., Kharybin, O. N., Antipin, M. Y. & Kostyanovsky, R. G. Lamellar racemic twinning as an obstacle for the resolution of enantiomers by crystallization: the case of Me(All)N⁺(CH₂Ph)Ph X⁻ (X = Br, I) salts. *J. Phys. Chem. B* **107**, 13523–13531 (2003).

21. 21.

Tanaka, K., Okada, T. & Toda, F. Separation of enantiomers of 2,2'-dihydroxy-1,1'-binaphthyl and 10,10'-dihydroxy-9,9'-biphenanthryl by complexation with *n*-alkylcinchonidinium halides. *Angew. Chem. Int. Edn Engl.* **32**, 1147–1148 (1993).

22. 22.

Toda, F., Tanaka, K., Stein, Z. & Goldberg, I. Optical resolution of binaphthyl and biphenanthryl diols by inclusion crystallization with *N*-alkylcinchonidium halides. Structural characterization of the resolved materials. *J. Org. Chem.* **59**, 5748–5751 (1994).

23. 23.

Du, H. et al. A new method for optical resolution of BINOL by molecular complexation with (*S*)-5-oxopyrrolidine-2-carboxanilide. *Tetrahedr. Lett.* **43**, 5273–5276 (2002).

24. 24.

Deng, J. et al. Resolution of omeprazole by inclusion complexation with a chiral host BINOL. *Tetrahedron Asymmetry* **11**, 1729–1732 (2000).

25. 25.

Schanz, H. J., Linseis, M. A. & Gilheany, D. G. Improved resolution methods for (*R,R*)- and (*S,S*)-cyclohexane-1,2-diamine and (*R*)- and (*S*)-BINOL. *Tetrahedron Asymmetry* **14**, 2763–2769 (2003).

26. 26.

Roy, B. N. et al. A novel method for large-scale synthesis of lamivudine through cocrystal formation of racemic lamivudine with (*S*)-(-)-1,1'-Bi(2-naphthol) [(*S*)-(BINOL)]. *Org. Process Res. Dev.* **13**, 450–455 (2009).

27. 27.

Ratajczak-Sitarz, M., Katrusiak, A., Gawrońska, K. & Gawroński, J. Racemate resolution via diastereomeric helicates in hydrogen-bonded co-crystals: the case of BINOL-diamine complexes. *Tetrahedron Asymmetry* **18**, 765–773 (2007).

28. 28.

Jin, S., Dong, Q., Wang, D. & Zhou, W. Six hydrogen bond directed supramolecular adducts formed between racemic-bis- β -naphthol and N-containing aromatic bases. *J. Mol. Struct.* **1013**, 143–155 (2012).

29. 29.

Tayama, E. & Tanaka, H. An efficient optical resolution of nitrogen-centered chiral beta-hydroxy-tetraalkylammonium salts via complexation with (*R*)-BINOL. *Tetrahedr. Lett.* **48**, 4183–4185 (2007).

30. 30.

Tayama, E., Otoyama, S. & Tanaka, H. Resolution of nitrogen-centered chiral tetraalkylammonium salts: application to [1,2] Stevens rearrangements with *N*-to-*C* chirality transmission. *Tetrahedron Asymmetry* **20**, 2600–2608 (2009).

31. 31.

Shirakawa, S. et al. Tetraalkylammonium salts as hydrogen-bonding catalysts. *Angew. Chem. Int. Ed.* **54**, 15767–15770 (2015).

32. 32.

Pike, S. J., Lavagnini, E., Varley, L. M., Cook, J. L. & Hunter, C. A. H-bond donor parameters for cations. *Chem. Sci.* **10**, 5943–5951 (2019).

33. 33.

Hunter, C. A. Quantifying intermolecular interactions: guidelines for the molecular recognition toolbox. *Angew. Chem. Int. Ed.* **43**, 5310–5324 (2004).

34. 34.

Taylor, R. & Kennard, O. Crystallographic evidence for the existence of C—H \cdots O, C—H \cdots N, and C—H \cdots Cl hydrogen bonds. *J. Am. Chem. Soc.* **104**, 5063–5070 (1982).

35. 35.

Lacour, J., Vial, L. & Herse, C. Efficient NMR enantiodifferentiation of chiral quats with BINPHAT anion. *Org. Lett.* **4**, 1351–1354 (2002).

36. 36.

Lacour, J., Londez, A., Goujon-Ginglinger, C., Buss, V. & Bernardinelli, G. Configurational ordering of cationic chiral dyes using a novel $C_{(2)}$ -symmetric hexacoordinated phosphate anion. *Org. Lett.* **2**, 4185–4188 (2000).

37. 37.

Michon, C., Gonçalves-Farbos, M.-H. & Lacour, J. NMR enantiodifferentiation of quaternary ammonium salts of Tröger base. *Chirality* **21**, 809–817 (2009).

38. 38.

Steed, K. M. & Steed, J. W. Packing problems: high Z crystal structures and their relationship to cocrystals, inclusion compounds, and polymorphism. *Chem. Rev.* **115**, 2895–2933 (2015).

39. 39.

Spackman, M. A. & Jayatilaka, D. Hirshfield surface analysis. *CrystEngComm* **11**, 19–32 (2009).

40. 40.

Gavezzotti, A. Are crystal structures predictable? *Acc. Chem. Res.* **27**, 309–314 (1994).

41. 41.

Gavezzotti, A. & Fillipini, G. Geometry of the intermolecular X–H \cdots Y (X, Y = N, O) hydrogen bond and the calibration of empirical hydrogen-bond potentials. *J. Phys. Chem.* **98**, 4831–4837 (1994).

42. 42.

Mullin, J. W. in *Ullman's Encyclopedia of Industrial Chemistry* Vol. 10 (ed. Elvers, B.) 582–630 (Wiley-VCH, 2012).

43. 43.

Kulchat, S. & Lehn, J.-M. Dynamic covalent chemistry of nucleophilic substitution component exchange of quaternary ammonium salts. *Chem. Asian J.* **10**, 2484–2496 (2015).

44. 44.

Lee, I., Park, Y. K., Huh, C. & Lee, H. W. Nucleophilic substitution reaction of benzyl bromide with *N,N*-dimethylaniline: significance of equilibrium cross-interaction constant. *J. Phys. Org. Chem.* **7**, 555–560 (1994).

45. 45.

Bordwell, F. G. & Hughes, D. L. Rate-equilibrium relationships for reactions of families of carbanion nucleophiles with *N*-benzyl-*N,N*-dimethylanilinium cations and with alkyl chlorides, bromides, and iodides. *J. Am. Chem. Soc.* **108**, 7300–7309 (1986).

46. 46.

Abboud, J.-L. M., Notario, R., Bertran, J. & Sold, M. in *Progress in Physical Organic Chemistry* (ed. Taft, R. W.) (Wiley, 1993).

47. 47.

Keith, J. M., Larwo, J. F. & Jacobsen, E. N. Practical considerations in kinetic resolution reactions. *Adv. Synth. Catal.* **343**, 5–26 (2001).

48. 48.

Brands, K. M. J. & Davies, A. J. Crystallization-induced diastereomer transformations. *Chem. Rev.* **106**, 2711–2733 (2006).

49. 49.

Flack, H. D. On enantiomorph-polarity estimation. *Acta Crystallogr. A* **39**, 876–881 (1983).

50. 50.

Genov, G. R., Douthwaite, J. L., Lahdenperä, A. S. K., Gibson, D. C. & Phipps, R. J. Enantioselective remote C–H activation directed by a chiral cation. *Science* **367**, 1246–1251 (2020).

51. 51.

Wang, Y., Sun, J. & Ding, K. Practical method and novel mechanism for optical resolution of BINOL by molecular complexation with *N*-benzylcinchoninium chloride. *Tetrahedron* **56**, 4447–4451 (2000).

Acknowledgements

We acknowledge funding from The Royal Society to M.O.K in the form of a University Research Fellowship (UF150536) and equipment grant (RGS\R2\180467) award. Durham University is acknowledged for providing a doctoral studentship (M.P.W.). J.M.P acknowledges support from the Laidlaw Undergraduate Research and Leadership programme in the form of a scholarship. The Royal Society is also acknowledged for providing M.E.L. with funding for a summer studentship. We thank our colleagues at Durham University and beyond, as well as members of the Kitching group, for input and advice during the preparation of this manuscript.

Author information

Affiliations

1. Department of Chemistry, Durham University, Durham, UK

Mark P. Walsh, Joseph M. Phelps, Marc E. Lennon, Dmitry S. Yufit & Matthew O. Kitching

Contributions

The project was conceived by M.O.K. and M.P.W. Experiments were devised by M.O.K. and M.P.W. M.P.W., J.M.P. and M.E.L. carried out starting material synthesis for the project. M.P.W. carried out experimental work to develop the enantioselective recognition, dynamic studies and enantioselective syntheses. X-ray crystallography was conducted by M.P.W. and D.S.Y. The manuscript was prepared by M.O.K. and M.P.W. with input from all authors.

Corresponding author

Correspondence to [Matthew O. Kitching](#).

Ethics declarations

Competing interests

The authors have filed a patent on this work (GB2017799.4).

Additional information

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Recognition screening.](#)

Addition of recognition species (0.5 equiv) to 60 mM solution (CDCl_3) of (*rac*)-**1b**. Recognition was monitored by observing changes in chemical shift and increased multiplicities of ^1H resonances of salt (*rac*)-**1b**.

Extended Data Fig. 2 Crystal structure of ternary complex **2b**.

a, The asymmetric unit ($P4_3$). **b**, Viewed along the *b* axis. **c**, Viewed along the *c* axis.

Extended Data Fig. 3 BINOL–halide network.

The (*R*)-BINOL and bromide counterions of complex **2d** are shown as a van der Waals surface (teal), displaying the chiral hydrogen-bond network that encapsulates the ammonium cation (*S*)-**1d**.

Extended Data Fig. 4 Control reactions.

a, Table of control reactions, demonstrating the requirement for correct balance of temperature, alkylating agent and concentration for optimal results. **b**, Analysis of both the solid and solution phases of the reaction mixture. Both phases show bias towards the (*S*) enantiomer of the quaternary ammonium cation.

Extended Data Fig. 5 Ammonium hexafluorophosphate salts.

a, X-ray crystal structures of enantioenriched hexafluorophosphate salts (*S*)-**1t** and (*R*)-**1t**. **b**, Evaluation of the stereochemical stability of (*S*)-**1t** and (*R*)-**1t** by exposing both enantiomers to conditions previously used to racemize ammonium halide salts, while also observing minimal changes to their optical activity after 24 h.

Supplementary information

Supplementary Information.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Walsh, M.P., Phelps, J.M., Lennon, M.E. *et al.* Enantioselective synthesis of ammonium cations. *Nature* **597**, 70–76 (2021).

<https://doi.org/10.1038/s41586-021-03735-5>

- Received: 01 March 2020
- Accepted: 17 June 2021
- Published: 01 September 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03735-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03735-5>

- Article
- [Published: 01 September 2021](#)

The contribution of insects to global forest deadwood decomposition

- [Sebastian Seibold](#) ORCID: orcid.org/0000-0002-7968-4489^{1,2,5,68},
- [Werner Rammer](#)¹,
- [Torsten Hothorn](#)³,
- [Rupert Seidl](#) ORCID: orcid.org/0000-0002-3338-3402^{1,2},
- [Michael D. Ulyshen](#)⁴,
- [Janina Lorz](#)⁵,
- [Marc W. Cadotte](#)⁶,
- [David B. Lindenmayer](#) ORCID: orcid.org/0000-0002-4766-4088⁷,
- [Yagya P. Adhikari](#)^{8,9},
- [Roxana Aragón](#)¹⁰,
- [Soyeon Bae](#) ORCID: orcid.org/0000-0003-1961-1226¹¹,
- [Petr Baldrian](#) ORCID: orcid.org/0000-0002-8983-2721¹²,
- [Hassan Barimani Varandi](#)¹³,
- [Jos Barlow](#) ORCID: orcid.org/0000-0003-4992-2594^{14,15},
- [Claus Bässler](#)^{16,17},
- [Jacques Beauchêne](#)¹⁸,
- [Erika Berenguer](#) ORCID: orcid.org/0000-0001-8157-8792^{14,19},
- [Rodrigo S. Bergamin](#)²⁰,
- [Tone Birkemoe](#)²¹,
- [Gergely Boros](#)^{22,23},
- [Roland Brandl](#)²⁴,
- [Hervé Brustel](#)²⁵,
- [Philip J. Burton](#) ORCID: orcid.org/0000-0002-5956-2716²⁶,

- [Yvonne T. Cakpo-Tossou²⁷](#),
- [Jorge Castro²⁸](#),
- [Eugénie Cateau^{25,29}](#),
- [Tyler P. Cobb³⁰](#),
- [Nina Farwig³¹](#),
- [Romina D. Fernández¹⁰](#),
- [Jennifer Firn ORCID: orcid.org/0000-0001-6026-8912^{32,33}](#),
- [Kee Seng Gan³⁴](#),
- [Grizelle González ORCID: orcid.org/0000-0003-3007-5540³⁵](#),
- [Martin M. Gossner ORCID: orcid.org/0000-0003-1516-6364³⁶](#),
- [Jan C. Habel³⁷](#),
- [Christian Hébert³⁸](#),
- [Christoph Heibl ORCID: orcid.org/0000-0002-7655-3299¹⁷](#),
- [Osmo Heikkala³⁹](#),
- [Andreas Hemp⁴⁰](#),
- [Claudia Hemp⁴⁰](#),
- [Joakim Hjältén⁴¹](#),
- [Stefan Hotes⁴²](#),
- [Jari Kouki ORCID: orcid.org/0000-0003-2624-8592⁴³](#),
- [Thibault Lachat ORCID: orcid.org/0000-0003-3952-7443^{36,44}](#),
- [Jie Liu⁴⁵](#),
- [Yu Liu ORCID: orcid.org/0000-0001-9869-2735⁴⁶](#),
- [Ya-Huang Luo⁴⁵](#),
- [Damasa M. Macandog ORCID: orcid.org/0000-0002-3700-1176⁴⁷](#),
- [Pablo E. Martina⁴⁸](#),
- [Sharif A. Mukul ORCID: orcid.org/0000-0001-6955-2469⁴⁹](#),
- [Baatarbileg Nachin⁵⁰](#),
- [Kurtis Nisbet ORCID: orcid.org/0000-0002-3536-7318⁵¹](#),
- [John O'Halloran⁵²](#),
- [Anne Oxbrough⁵³](#),
- [Jeev Nath Pandey⁵⁴](#),
- [Tomáš Pavlíček⁵⁵](#),
- [Stephen M. Pawson^{56,57}](#),
- [Jacques S. Rakotondranary^{58,59}](#),

- [Jean-Baptiste Ramanamanjato⁶⁰](#),
- [Liana Rossi⁶¹](#),
- [Jürgen Schmidl](#) ORCID: orcid.org/0000-0001-5497-4116⁶²,
- [Mark Schulze](#) ORCID: orcid.org/0000-0003-0438-4091⁶³,
- [Stephen Seaton](#) ORCID: orcid.org/0000-0002-1076-6005⁶⁴,
- [Marisa J. Stone](#) ORCID: orcid.org/0000-0002-5766-0646⁶⁵,
- [Nigel E. Stork⁶⁵](#),
- [Byambagerel Suran](#) ORCID: orcid.org/0000-0002-7078-5566⁵⁰,
- [Anne Sverdrup-Thygeson](#) ORCID: orcid.org/0000-0002-3122-2250²¹,
- [Simon Thorn⁵](#),
- [Ganesh Thyagarajan⁶⁶](#),
- [Timothy J. Wardlaw](#) ORCID: orcid.org/0000-0002-8686-0671⁶⁷,
- [Wolfgang W. Weisser](#) ORCID: orcid.org/0000-0002-2757-8959⁶⁸,
- [Sungsoo Yoon](#) ORCID: orcid.org/0000-0003-4712-8004⁶⁹,
- [Naili Zhang⁷⁰](#) &
- [Jörg Müller](#) ORCID: orcid.org/0000-0002-1409-1586^{5,17}

Nature volume 597, pages 77–81 (2021)

- 1978 Accesses
- 333 Altmetric
- [Metrics details](#)

Subjects

- [Biodiversity](#)
- [Climate and Earth system modelling](#)
- [Ecosystem ecology](#)
- [Forest ecology](#)

Abstract

The amount of carbon stored in deadwood is equivalent to about 8 per cent of the global forest carbon stocks¹. The decomposition of deadwood is largely governed by climate^{2,3,4,5} with decomposer groups—such as microorganisms and insects—contributing to variations in the decomposition rates^{2,6,7}. At the global scale, the contribution of insects to the decomposition of deadwood and carbon release remains poorly understood⁷. Here we present a field experiment of wood decomposition across 55 forest sites and 6 continents. We find that the deadwood decomposition rates increase with temperature, and the strongest temperature effect is found at high precipitation levels. Precipitation affects the decomposition rates negatively at low temperatures and positively at high temperatures. As a net effect—including the direct consumption by insects and indirect effects through interactions with microorganisms— insects accelerate the decomposition in tropical forests (3.9% median mass loss per year). In temperate and boreal forests, we find weak positive and negative effects with a median mass loss of 0.9 per cent and −0.1 per cent per year, respectively. Furthermore, we apply the experimentally derived decomposition function to a global map of deadwood carbon synthesized from empirical and remote-sensing data, obtaining an estimate of 10.9 ± 3.2 petagram of carbon per year released from deadwood globally, with 93 per cent originating from tropical forests. Globally, the net effect of insects may account for 29 per cent of the carbon flux from deadwood, which suggests a functional importance of insects in the decomposition of deadwood and the carbon cycle.

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: Decomposition rates and insect effects per biome.

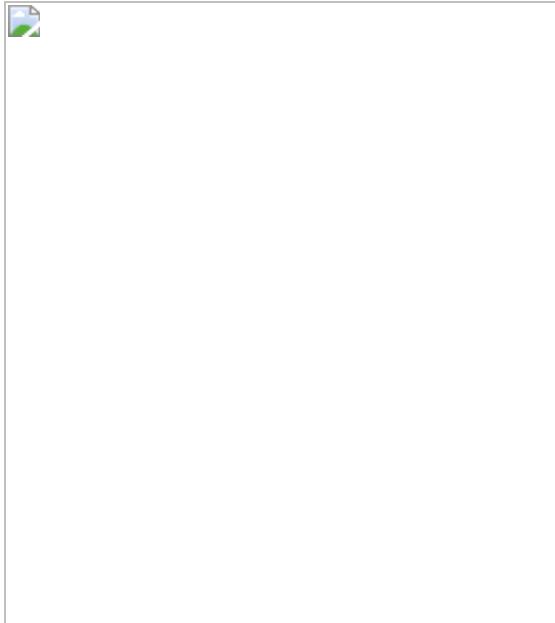


Fig. 2: Decomposition rates and net insect effects in climate space.

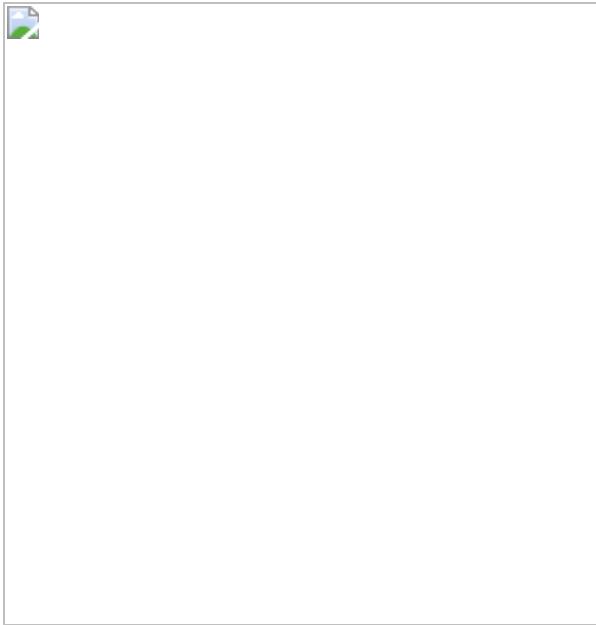


Fig. 3: Global annual carbon release from deadwood and sensitivity analysis.



Data availability

Raw data from the global deadwood experiment, our global map of deadwood carbon and our map of predicted decomposition rates are publicly available from Figshare

<https://figshare.com/s/ffc39ee0724b11bf450c>
[\(https://doi.org/10.6084/m9.figshare.14545992\)](https://doi.org/10.6084/m9.figshare.14545992).

Code availability

An annotated R code including the data needed to reproduce the statistical analyses, global estimates and sensitivity analysis is publicly available from

Figshare <https://figshare.com/s/ffc39ee0724b11bf450c>
(<https://doi.org/10.6084/m9.figshare.14545992>).

References

1. 1.
Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
2. 2.
Bradford, M. A. et al. Climate fails to predict wood decomposition at regional scales. *Nat. Clim. Change* **4**, 625–630 (2014).
3. 3.
Chambers, J. Q., Higuchi, N., Schimel, J. P. J., Ferreira, L. V. & Melack, J. M. Decomposition and carbon cycling of dead trees in tropical forests of the central Amazon. *Oecologia* **122**, 380–388 (2000).
4. 4.
González, G. et al. Decay of aspen (*Populus tremuloides* Michx.) wood in moist and dry boreal, temperate, and tropical forest fragments. *Ambio* **37**, 588–597 (2008).
5. 5.
Stokland, J., Siitonen, J. & Jonsson, B. G. *Biodiversity in Dead Wood* (Cambridge Univ. Press, 2012).
6. 6.
Lustenhouwer, N. et al. A trait-based understanding of wood decomposition by fungi. *Proc. Natl Acad. Sci. USA* **117**, 11551–11558 (2020).

7. 7.

Ulyshen, M. D. Wood decomposition as influenced by invertebrates. *Biol. Rev. Camb. Philos. Soc.* **91**, 70–85 (2016).

8. 8.

Pretzsch, H., Biber, P., Schütze, G., Uhl, E. & Rötzer, T. Forest stand growth dynamics in Central Europe have accelerated since 1870. *Nat. Commun.* **5**, 4967 (2014).

9. 9.

Büntgen, U. et al. Limited capacity of tree growth to mitigate the global greenhouse effect under predicted warming. *Nat. Commun.* **10**, 2171 (2019).

10. 10.

Seidl, R. et al. Forest disturbances under climate change. *Nat. Clim. Change* **7**, 395–402 (2017).

11. 11.

Hubau, W. et al. Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* **579**, 80–87 (2020).

12. 12.

Portillo-Estrada, M. et al. Climatic controls on leaf litter decomposition across European forests and grasslands revealed by reciprocal litter transplantation experiments. *Biogeosciences* **13**, 1621–1633 (2016).

13. 13.

Christenson, L. et al. Winter climate change influences on soil faunal distribution and abundance: implications for decomposition in the northern forest. *Northeast. Nat.* **24**, B209–B234 (2017).

14. 14.

Keenan, T. F. et al. Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499**, 324–327 (2013).

15. 15.

Stephenson, N. L. et al. Rate of tree carbon accumulation increases continuously with tree size. *Nature* **507**, 90–93 (2014).

16. 16.

Martin, A., Dimke, G., Doraisami, M. & Thomas, S. Carbon fractions in the world's dead wood. *Nat. Commun.* **12**, 889 (2021).

17. 17.

Friedlingstein, P. et al. Global carbon budget 2019. *Earth Syst. Sci. Data* **11**, 1783–1838 (2019).

18. 18.

Marshall, D. J., Pettersen, A. K., Bode, M. & White, C. R. Developmental cost theory predicts thermal environment and vulnerability to global warming. *Nat. Ecol. Evol.* **4**, 406–411 (2020).

19. 19.

Buczkowski, G. & Bertelsmeier, C. Invasive termites in a changing climate: a global perspective. *Ecol. Evol.* **7**, 974–985 (2017).

20. 20.

Diaz, S., Settele, J. & Brondizio, E. *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES, 2019).

21. 21.

van Klink, R. et al. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* **368**, 417–420 (2020).

22. 22.

Seibold, S. et al. Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature* **574**, 671–674 (2019).

23. 23.

Harris, N. L. et al. Global maps of twenty-first century forest carbon fluxes. *Nat. Clim. Change* **11**, 234–240 (2021).

24. 24.

Jacobsen, R. M., Sverdrup-Thygeson, A., Kauserud, H., Mundra, S. & Birkemoe, T. Exclusion of invertebrates influences saprotrophic fungal community and wood decay rate in an experimental field study. *Funct. Ecol.* **32**, 2571–2582 (2018).

25. 25.

Skelton, J. et al. Fungal symbionts of bark and ambrosia beetles can suppress decomposition of pine sapwood by competing with wood-decay fungi. *Fungal Ecol.* **45**, 100926 (2020).

26. 26.

Wu, D., Seibold, S., Ruan, Z., Weng, C. & Yu, M. Island size affects wood decomposition by changing decomposer distribution. *Ecography* **44**, 456–468 (2021).

27. 27.

Harmon, M. E. et al. Release of coarse woody detritus-related carbon: a synthesis across forest biomes. *Carbon Balance Manag.* **15**, 1

(2020).

28. 28.

Wall, D. H. et al. Global decomposition experiment shows soil animal impacts on decomposition are climate-dependent. *Glob. Change Biol.* **14**, 2661–2677 (2008).

29. 29.

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293**, 2248–2251 (2001).

30. 30.

Baldrian, P. et al. Responses of the extracellular enzyme activities in hardwood forest to soil temperature and seasonality and the potential effects of climate change. *Soil Biol. Biochem.* **56**, 60–68 (2013).

31. 31.

A'Bear, A. D., Jones, T. H., Kandeler, E. & Boddy, L. Interactive effects of temperature and soil moisture on fungal-mediated wood decomposition and extracellular enzyme activity. *Soil Biol. Biochem.* **70**, 151–158 (2014).

32. 32.

IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* (IPCC, 2014).

33. 33.

Smyth, C. E., Kurz, W. A., Trofymow, J. A. & CIDET Working Group. Including the effects of water stress on decomposition in the Carbon Budget Model of the Canadian Forest Sector CBM-CFS3. *Ecol. Modell.* **222**, 1080–1091 (2011).

34. 34.

Weedon, J. T. et al. Global meta-analysis of wood decomposition rates: a role for trait variation among tree species? *Ecol. Lett.* **12**, 45–56 (2009).

35. 35.

Griffiths, H. M., Ashton, L. A., Evans, T. A., Parr, C. L. & Eggleton, P. Termites can decompose more than half of deadwood in tropical rainforest. *Curr. Biol.* **29**, R118–R119 (2019).

36. 36.

Birkemoe, T., Jacobsen, R. M., Sverdrup-Thygeson, A. & Biedermann, P. H. W. in *Saproxylic Insects* (ed. Ulyshen, M. D.) 377–427 (Springer, 2018).

37. 37.

Harvell, M. C. E. et al. Climate warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158–2162 (2002).

38. 38.

Berkov, A. in *Saproxylic Insects* (ed. Ulyshen, M. D.) 547–580 (Springer, 2018).

39. 39.

Wang, C., Bond-Lamberty, B. & Gower, S. T. Environmental controls on carbon dioxide flux from black spruce coarse woody debris. *Oecologia* **132**, 374–381 (2002).

40. 40.

Peršoh, D. & Borken, W. Impact of woody debris of different tree species on the microbial activity and community of an underlying organic horizon. *Soil Biol. Biochem.* **115**, 516–525 (2017).

41. 41.

Campbell, J., Donato, D., Azuma, D. & Law, B. Pyrogenic carbon emission from a large wildfire in Oregon, United States. *J. Geophys. Res.* **112**, G04014 (2007).

42. 42.

van Leeuwen, T. T. et al. Biomass burning fuel consumption rates: a field measurement database. *Biogeosciences* **11**, 7305–7329 (2014).

43. 43.

McDowell, N. G. et al. Pervasive shifts in forest dynamics in a changing world. *Science* **368**, eaaz9463 (2020).

44. 44.

Ulyshen, M. D. & Wagner, T. L. Quantifying arthropod contributions to wood decay. *Methods Ecol. Evol.* **4**, 345–352 (2013).

45. 45.

Bässler, C., Heilmann-Clausen, J., Karasch, P., Brandl, R. & Halbwachs, H. Ectomycorrhizal fungi have larger fruit bodies than saprotrophic fungi. *Fungal Ecol.* **17**, 205–212 (2015).

46. 46.

Ryvarden, L. & Gilbertson, R. L. *The Polyporaceae of Europe* (Fungiflora, 1994).

47. 47.

Eriksson, J. & Ryvarden, L. *The Corticiaceae of North Europe* Parts 1–8 (Fungiflora, 1987).

48. 48.

Boddy, L., Hynes, J., Bebber, D. P. & Fricker, M. D. Saprotrophic cord systems: dispersal mechanisms in space and time. *Mycoscience* **50**, 9–19 (2009).

49. 49.

Moore, D. *Fungal Morphogenesis* (Cambridge Univ. Press, 1998).

50. 50.

Clemenccon, H. *Anatomy of the Hymenomycetes* (Univ. Lausanne, 1997).

51. 51.

R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2020).

52. 52.

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).

53. 53.

Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

54. 54.

Wood, S. N. *Generalized Additive Models: an Introduction with R* 2nd edn (Chapman and Hall/CRC, 2017).

55. 55.

Robinson, D. Implications of a large global root biomass for carbon sink estimates and for soil carbon dynamics. *Proc. R. Soc. B* **274**, 2753–2759 (2007).

56. 56.

Food and Agriculture Organization. *Global Ecological Zones for FAO Forest Reporting: 2010 Update, Forest Resource Assessment Working Paper* (Food and Agriculture Organization, 2012).

57. 57.

Food and Agriculture Organization. *Global Forest Resources Assessment 2015* (Food and Agriculture Organization, 2016).

58. 58.

Christensen, M. et al. Dead wood in European beech (*Fagus sylvatica*) forest reserves. *For. Eco. Man.* **210**, 267–282 (2005).

59. 59.

Kobayashi, T. et al. Production of global land cover data – GLCNMO2013. *J. Geogr. Geol.* **9**, 1–15 (2017).

60. 60.

Harmon, M. E., Woodall, C. W., Fasth, B., Sexton, J. & Yatkov, M. *Differences between Standing and Downed Dead Tree Wood Density Reduction Factors: A Comparison across Decay Classes and Tree Species* Research Paper NRS-15 (US Department of Agriculture, Forest Service, Northern Research Station, 2011).

61. 61.

Hararuk, O., Kurz, W. A. & Didion, M. Dynamics of dead wood decay in Swiss forests. *For. Ecosyst.* **7**, 36 (2020).

62. 62.

Gora, E. M., Kneale, R. C., Larjavaara, M. & Muller-Landau, H. C. Dead wood necromass in a moist tropical forest: stocks, fluxes, and spatiotemporal variability. *Ecosystems* **22**, 1189–1205 (2019).

63. 63.

Héault, B. et al. Modeling decay rates of dead wood in a neotropical forest. *Oecologia* **164**, 243–251 (2010).

64. 64.

Thünen-Institut für Waldökosysteme. *Der Wald in Deutschland - Ausgewählte Ergebnisse der dritten Bundeswaldinventur* (Bundesministerium für Ernährung und Landwirtschaft, 2014).

65. 65.

Puletti, N. et al. A dataset of forest volume deadwood estimates for Europe. *Ann. For. Sci.* **76**, 68 (2019).

66. 66.

Richardson, S. J. et al. Deadwood in New Zealand's indigenous forests. *For. Ecol. Manage.* **258**, 2456–2466 (2009).

67. 67.

Shorohova, E. & Kapitsa, E. Stand and landscape scale variability in the amount and diversity of coarse woody debris in primeval European boreal forests. *For. Ecol. Manage.* **356**, 273–284 (2015).

68. 68.

Szymański, C., Fontana, G. & Sanguinetti, J. Natural and anthropogenic influences on coarse woody debris stocks in Nothofagus–Araucaria forests of northern Patagonia, Argentina. *Austral Ecol.* **42**, 48–60 (2017).

69. 69.

Link, K. G. et al. A local and global sensitivity analysis of a mathematical model of coagulation and platelet deposition under flow. *PLoS One* **13**, e0200917 (2018).

70. 70.

Saugier, B., Roy, J. & Mooney, H. A. in *Terrestrial Global Productivity* (eds J. Roy, B. Saugier & H. A. Mooney) 543–557 (Academic Press, 2001).

Acknowledgements

We thank the administration of the Bavarian Forest National Park for financing the set-up of the experiment and all members of the local teams for their contribution in the field and laboratory. We especially thank D. Blair who operated the site in Victoria, Australia, until his unexpected death in 2019. We thank B. von Rentzel, J. Ganzhorn, A. Gruppe, M. Harmon, S. Muller and S. Irwin, Makiling Center for Mountain Ecosystems, University of the Philippines Los Banos, the Ministerio del Ambiente de Ecuador, the Instituto Nacional de Biodiversidad de Ecuador and the foundation “Nature and Culture International” for their support. S. Seibold was supported by the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research and the People Programme of the European Union (Marie Curie Actions; grant number 605728). N.F. was supported by the German Research Foundation (FA925/7-1, FA925/11-1).

Author information

Affiliations

1. Ecosystem Dynamics and Forest Management Group, School of Life Sciences, Technical University of Munich, Freising, Germany

Sebastian Seibold, Werner Rammer & Rupert Seidl

2. Berchtesgaden National Park, Berchtesgaden, Germany

Sebastian Seibold & Rupert Seidl

3. Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Torsten Hothorn

4. Southern Research Station, USDA Forest Service, Athens, GA, USA

Michael D. Ulyshen

5. Field Station Fabrikschleichach, University of Würzburg, Rauhenebrach, Germany

Sebastian Seibold, Janina Lorz, Simon Thorn & Jörg Müller

6. Biological Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

Marc W. Cadotte

7. Fenner School of Environment and Society, The Australian National University, Canberra, Australian Capital Territory, Australia

David B. Lindenmayer

8. Department of Biogeography, University of Bayreuth, Bayreuth, Germany

Yagya P. Adhikari

9. Department of Disturbance Ecology, University of Bayreuth, Bayreuth, Germany

Yagya P. Adhikari

10. Instituto de Ecología Regional, CONICET-Universidad Nacional de Tucumán, Yerba Buena, Argentina

Roxana Aragón & Romina D. Fernández

11. Department of Animal Ecology and Tropical Biology, University of Würzburg, Würzburg, Germany

Soyeon Bae

12. Laboratory of Environmental Microbiology, Institute of Microbiology, The Czech Academy of Sciences, Prague, Czech Republic

Petr Baldrian

13. Agricultural and Natural Resources Research Centre of Mazandaran, Sari, Iran

Hassan Barimani Varandi

14. Lancaster Environment Centre, Lancaster University, Lancaster, UK

Jos Barlow & Erika Berenguer

15. Universidade Federal de Lavras, Lavras, Brazil

Jos Barlow

16. Department of Biodiversity Conservation, Goethe-University Frankfurt, Frankfurt, Germany

Claus Bässler

17. Bavarian Forest National Park, Grafenau, Germany

Claus Bässler, Christoph Heibl & Jörg Müller

18. CIRAD, UMR Ecologie des Forêts de Guyane (EcoFoG), AgroParisTech, CNRS, INRA, Universite des Antilles, Universite de Guyane, Kourou, France

Jacques Beauchêne

19. Environmental Change Institute, University of Oxford, Oxford, UK

Erika Berenguer

20. Grassland Vegetation Lab, Federal University of Rio Grande do Sul,
Porto Alegre, Brazil

Rodrigo S. Bergamin

21. Faculty of Environmental Sciences and Natural Resource
Management, Norwegian University of Life Sciences, Aas, Norway

Tone Birkemoe & Anne Sverdrup-Thygeson

22. Institute of Ecology and Botany, Centre for Ecological Research,
Vácrátót, Hungary

Gergely Boros

23. Institute for Wildlife Management and Nature Conservation,
Hungarian University of Agriculture and Life Sciences, Gödöllő,
Hungary

Gergely Boros

24. Animal Ecology, University of Marburg, Marburg, Germany

Roland Brandl

25. École d'Ingénieurs de Purpan, Université de Toulouse, UMR 1201
Dynafor, Toulouse, France

Hervé Brustel & Eugénie Cateau

26. Ecosystem Science and Management Program, University of Northern
British Columbia, Terrace, British Columbia, Canada

Philip J. Burton

27. Laboratory of Applied Ecology, University of Abomey-Calavi,
Godomey, Benin

Yvonne T. Cakpo-Tossou

28. Department of Ecology, University of Granada, Granada, Spain

Jorge Castro

29. Réserves Naturelles de France, Dijon, France

Eugénie Cateau

30. Royal Alberta Museum, Edmonton, Alberta, Canada

Tyler P. Cobb

31. Conservation Ecology, University of Marburg, Marburg, Germany

Nina Farwig

32. Science and Engineering Faculty, Queensland University of Technology, Brisbane, Queensland, Australia

Jennifer Firn

33. Centre for the Environment, Institute for Future Environments, Brisbane, Queensland, Australia

Jennifer Firn

34. Forest Research Institute Malaysia, Kuala Lumpur, Malaysia

Kee Seng Gan

35. International Institute of Tropical Forestry, USDA Forest Service, San Juan, PR, USA

Grizelle González

36. Forest Entomology, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

Martin M. Gossner & Thibault Lachat

37. Evolutionary Zoology, University of Salzburg, Salzburg, Austria

Jan C. Habel

38. Natural Resources Canada, Canadian Forest Service, Quebec, Quebec, Canada

Christian Hébert

39. Eurofins Ahma Oy, Oulu, Finland

Osmo Heikkala

40. Department of Plant Systematics, University of Bayreuth, Bayreuth, Germany

Andreas Hemp & Claudia Hemp

41. Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden

Joakim Hjältén

42. Applied Landscape Ecology, Chuo University, Tokyo, Japan

Stefan Hotes

43. School of Forest Sciences, University of Eastern Finland, Joensuu, Finland

Jari Kouki

44. School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, Zollikofen, Switzerland

Thibault Lachat

45. CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

Jie Liu & Ya-Huang Luo

46. ECNU-Alberta Joint Lab for Biodiversity Study, Tiantong National Station for Forest Ecosystem Research, East China Normal University, Shanghai, China

Yu Liu

47. Institute of Biological Sciences, University of the Philippines Los Banos, Laguna, The Philippines

Damasa M. Macandog

48. Department of Thermodynamics, Universidad Nacional del Nordeste, Resistencia, Argentina

Pablo E. Martina

49. Tropical Forests and People Research Centre, University of the Sunshine Coast, Maroochydore, Queensland, Australia

Sharif A. Mukul

50. Forest Ecosystem Monitoring Laboratory, National University of Mongolia, Ulaanbaatar, Mongolia

Baatarbileg Nachin & Byambagerel Suran

51. School of Environment and Science, Griffith University, Nathan, Queensland, Australia

Kurtis Nisbet

52. School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

John O'Halloran

53. Edge Hill University, Ormskirk, UK

Anne Oxbrough

54. Institute of Forestry, Tribhuvan University, Pokhara, Nepal

Jeev Nath Pandey

55. Institute of Evolution, University of Haifa, Haifa, Israel

Tomáš Pavláček

56. Scion (New Zealand Forest Research Institute), Christchurch, New Zealand

Stephen M. Pawson

57. School of Forestry, University of Canterbury, Christchurch, New Zealand

Stephen M. Pawson

58. Institute of Zoology, University of Hamburg, Hamburg, Germany

Jacques S. Rakotondranary

59. Faculté des Sciences, Université d'Antananarivo, Antananarivo, Madagascar

Jacques S. Rakotondranary

60. Tropical Biodiversity and Social Enterprise, Fort Dauphin, Madagascar

Jean-Baptiste Ramanamanjato

61. Departamento de Ecologia, Universidade Estadual Paulista, Rio Claro, Brazil

Liana Rossi

62. Ecology Group, University Erlangen-Nuremberg, Erlangen, Germany

Jürgen Schmidl

63. H. J. Andrews Experimental Forest, Blue River, OR, USA

Mark Schulze

64. Environmental and Conservation Sciences, Murdoch University, Melville, Western Australia, Australia

Stephen Seaton

65. Environmental Futures Research Institute, Griffith University, Nathan, Queensland, Australia

Marisa J. Stone & Nigel E. Stork

66. Ashoka Trust for Research in Ecology and the Environment, Bangalore, India

Ganesh Thyagarajan

67. ARC Centre for Forest Value, University of Tasmania, Hobart, Tasmania, Australia

Timothy J. Wardlaw

68. Terrestrial Ecology Research Group, School of Life Sciences, Technical University of Munich, Freising, Germany

Sebastian Seibold & Wolfgang W. Weisser

69. EcoBank Team, National Institute of Ecology, Seocheon-gun, Republic of Korea

Sungsoo Yoon

70. College of Forestry, Beijing Forestry University, Beijing, China

Naili Zhang

Contributions

S. Seibold, J.M. and R.S. conceived the idea of this manuscript. S. Seibold, J.M. and M.D.U. designed the experiment with inputs from P.B, C.B., R.B., M.M.G., J.S. and S.T. S. Seibold, J. Lorz, W.R., M.D.U., Y.P.A., R.A., S.B., H.B.V., J. Barlow, J. Beauchêne, E.B., R.S.B., T.B., G.B., H.B., P.J.B., M.W.C., Y.T.C.-T., J.C., E.C., T.P.C., N.F., R.D.F., J.F., K.S.G., G.G., J.C.H., C. Hébert, O.H., A.H., C. Hemp, J.H., S.H., J.K., T.L., D.B.L., J. Liu, Y.L., Y.-H.L., D.M.M., P.E.M., S.A.M., B.N., K.N., J.O'H., A.O., J.N.P., T.P., S.M.P., J.S.R., J.-B.R., L.R., M.S., S. Seaton, M.J.S., N.E.S., B.S., A.S.-T., G.T., T.J.W., S.Y., N.Z. and J.M. collected data. S. Seibold, T.H. and W.R. analysed the data. S. Seibold, J.M., R.S. and W.R. wrote the first manuscript draft with considerable inputs from M.D.U., M.W.C. and D.B.L. and finalized the manuscript. All authors commented on the manuscript.

Corresponding author

Correspondence to [Sebastian Seibold](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Robert M. Ewers 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 Arrangement of installations per site and per treatment.

a, Each site received three installations of three treatments randomly assigned to a 3×3 grid. **b–d**, Treatments included closed cages to exclude insects (**b**), open cages providing similar microclimatic conditions as closed cages but giving access to insects (**c**) and uncaged bundles of logs (**d**). Cages measured 40 cm \times 40 cm \times 60 cm and were made of white polyester with honeycomb-shaped meshes with a side length of approximately 0.5 mm. Open cages had four rectangular openings measuring 3 cm \times 12 cm at both front sides and four rectangular openings measuring 10 cm \times 15 cm at the bottom representing in total 6% of the surface area of the cage as well as a total of ten 12-cm slits at the top and long sides. All cages were placed on a stainless-steel mesh (0.5 mm mesh width), which had the same openings as the bottom side of the cages in the open-cage treatment. Photographs show the site in the Bavarian Forest National Park, Germany.

Extended Data Fig. 2 Effects of treatments on wood decomposition and insect colonization.

a, b, Coefficients and confidence intervals from post hoc tests assessing all three pairwise comparisons between the uncaged, closed-cage and open-cage treatments for annual mass loss (**a**; same structure as the model shown in Table 1 based on 3,578 logs) and insect colonization (**b**; binomial model for insect presence and absence based on 3,430 logs) of wood of native tree species. The 95% confidence intervals that do not intersect the zero line (dashed) indicate significant differences. **c**, Pairwise comparison of fitted

annual mass loss (%) between each of the three treatments in the global deadwood decomposition experiment. Points represent the predicted values for angiosperm species at 55 sites and gymnosperm species at 21 sites based on three Gaussian generalized linear mixed log-link models for 3,758 logs with site-specific random effects and temperature, precipitation, treatment (closed cage versus uncaged, open cage versus uncaged and closed cage versus open cage), host division, as well as their interactions, as fixed effects. In **a** and **b**, the largest differences in both response variables were observed between uncaged and closed-cage treatments. Annual mass loss was higher in the uncaged than open-cage treatment and higher in the open-cage than in closed-cage treatment, although the latter was not significant. This indicates that the open cage, despite its openings for insects, has a clearly reduced decomposition rate compared with the uncaged treatment. Insect colonization for the open cage differed significantly from both uncaged and closed-cage treatment, but was more similar to the uncaged than closed-cage treatment. This indicates that open cages were colonized by insects, but not as frequently as the uncaged treatment. Open cages thus excluded parts of the wood-decomposing insect community, which may explain the rather small difference in annual mass loss between closed cages and open cages. These results suggest that the comparison of uncaged wood versus closed cages provides a more reliable estimate of the net effect of insects on wood decomposition than the comparison of closed-cage versus open-cage treatments, which is likely to underestimate the net effect of insects. In **c**, the difference between annual mass loss in closed-cage and both treatments with insect access (uncaged and open cage) increased from boreal to tropical biomes, whereas the difference between uncaged wood and open cages hardly deviated from the 1:1 line. This indicates that the reported mass loss differences between closed-cage and uncaged treatments, as well as the accelerating effect of temperature and precipitation (Table 1), can be attributed to insects and are not an artefact of potential microclimatic effects of the cages (Supplementary Information section 1).

Extended Data Fig. 3 Interaction effects of temperature and precipitation on wood decomposition.

a, b, Predictions based on the model presented in Table 1 for annual mass loss of deadwood of native tree species (2,533 logs at 55 sites), considering all possible groups of decomposers (uncaged treatment) (**a**), and annual mass loss attributed to insects (difference in mass loss between uncaged and closed-cage treatments) (**b**), relative to temperature and precipitation. The length of the lines is limited to the gradients in precipitation covered by the sites.

Extended Data Fig. 4 Model evaluation against independent data.

Comparison of 157 independent observations of annual deadwood decomposition rates measured for larger diameter wood in previous deadwood surveys²⁷ (red dots) with the predictions from our model for the same locations (blue triangles). Lines indicate the relationship between the decomposition rate and mean annual temperature from Harmon et al.²⁷ (red dashed line; $k = 0.0184e^{0.0787 \times \text{temperature}}$) and for our model (blue line; $k = 0.0171e^{0.0812 \times \text{temperature}}$). Good correspondence of both curves indicates that our models of global carbon release from deadwood provide robust estimates despite being based on experimental deadwood with a diameter of around 3 cm (for detailed discussion, see Supplementary Information section 1).

Extended Data Fig. 5 Global deadwood carbon fluxes.

a, b, Total annual release of deadwood carbon from decomposition including all decomposers (**a**) and annual release of deadwood carbon due to the net effect of insects (**b**). Light grey areas indicate values of $\pm 0.1 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ and white areas are non-forest systems. **c**, Latitudinal distribution of global deadwood carbon fluxes per hectare.

Extended Data Fig. 6 Processing steps for the global deadwood carbon map.

a, Aboveground forest biomass (Mg ha^{-1}) aggregated to 5' from the GlobBiom dataset. **b**, Total live carbon (Mg ha^{-1}) by extending **a** with root

biomass⁵⁵ and conversion to carbon. **c**, Proportion of gymnosperm forests derived from the GLCNMO2013⁵⁹ dataset. The proportion of angiosperm cover is $1 - \text{gymnosperm cover}$. White indicates non-forested area.

Extended Data Fig. 7 Bioclimatic space for robust predictions.

a, b, Climate conditions outside of the range of prediction models for angiosperm (**a**) and gymnosperm (**b**) species in climate space (left) and mapped (right). Left, dark blue points are outside of the range defined by a convex hull around the experimental sites (black triangles). Right, the colours on the maps indicate the absolute difference between the local climate and the climate used for prediction for temperature (red colour channel) and precipitation (blue colour channel) with black indicating no difference. White areas indicate that no gymnosperm or angiosperm forest, respectively, occurs there. Experimental sites are indicated by yellow dots. Temperatures outside of the range are mainly located in northeastern Siberia and northern Canada, whereas offsets in precipitation are stronger for gymnosperms in southeastern Asia, Indonesia and in the Amazon region. The land surface area not covered by our experimental data is 23.5% for gymnosperms and 17.7% for angiosperms, representing together 13.2% of the carbon stored in deadwood. These areas were included in our upscaling by mapping them to the nearest point at the convex hull in climate space.

Extended Data Table 1 Supporting analyses of drivers of wood decomposition

Extended Data Table 2 Uncertainty in global carbon fluxes from the decomposition of deadwood, determined in a global sensitivity analysis

Extended Data Table 3 Comparison of global carbon stock estimates and results for each biome

Supplementary information

Supplementary Information

This Supplementary Information file contains the following sections: (1) methodological aspects of the wood decomposition experiment; (2)

challenges related to the upscaling from experimental data to global deadwood carbon fluxes; (3) overview of tree species and exposure time per site; and the Supplementary References.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Seibold, S., Rammer, W., Hothorn, T. *et al.* The contribution of insects to global forest deadwood decomposition. *Nature* **597**, 77–81 (2021).
<https://doi.org/10.1038/s41586-021-03740-8>

- Received: 07 June 2020
- Accepted: 18 June 2021
- Published: 01 September 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03740-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03740-8>

- Article
- [Published: 11 August 2021](#)

A metabolic function of the hippocampal sharp wave-ripple

- [David Tingley](#) ¹,
- [Kathryn McClain](#)²,
- [Ekin Kaya](#) [ORCID: orcid.org/0000-0001-9235-4954](#)^{1,3},
- [Jordan Carpenter](#)^{1,4} &
- [György Buzsáki](#) [ORCID: orcid.org/0000-0002-3100-4800](#)^{1,2,5,6}

[Nature](#) volume **597**, pages 82–86 (2021)

- 11k Accesses
- 1 Citations
- 569 Altmetric
- [Metrics details](#)

Subjects

- [Neuroendocrinology](#)
- [Neurophysiology](#)
- [Non-REM sleep](#)

Abstract

The hippocampus has previously been implicated in both cognitive and endocrine functions^{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}. We simultaneously

measured electrophysiological activity from the hippocampus and interstitial glucose concentrations in the body of freely behaving rats to identify an activity pattern that may link these disparate functions of the hippocampus. Here we report that clusters of sharp wave-ripples recorded from the hippocampus reliably predicted a decrease in peripheral glucose concentrations within about 10 min. This correlation was not dependent on circadian, ultradian or meal-triggered fluctuations, could be mimicked with optogenetically induced ripples in the hippocampus (but not in the parietal cortex) and was attenuated to chance levels by pharmacogenetically suppressing activity of the lateral septum, which is the major conduit between the hippocampus and the hypothalamus. Our findings demonstrate that a function of the sharp wave-ripple is to modulate peripheral glucose homeostasis, and offer a mechanism for the link between sleep disruption and blood glucose dysregulation in type 2 diabetes^{[16,17,18](#)}.

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: Chronic recording of interstitial glucose dynamics and CA1 in freely moving rats.

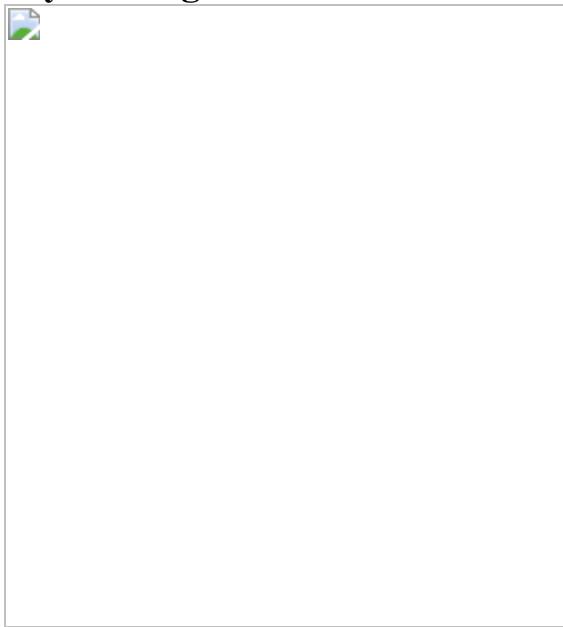


Fig. 2: SPW-R rate predicts future glucose fluctuations.

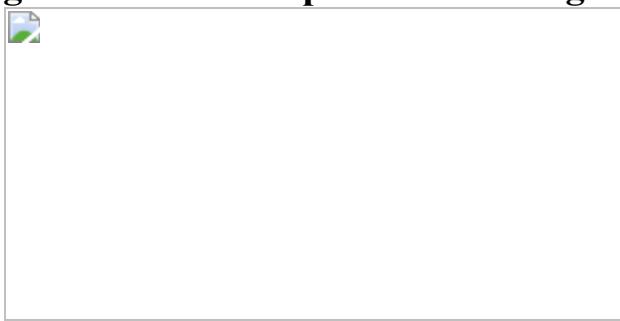


Fig. 3: Optogenetic induction of ripples mimics negative correlation with glucose fluctuations.

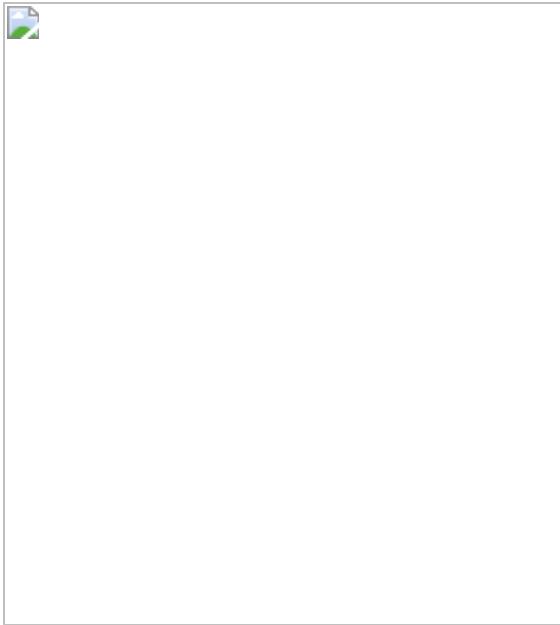


Fig. 4: Pharmacogenetic decoupling of SPW-Rs from glucose dynamics via the lateral septum.



Data availability

The datasets used during this study are available from
<https://buzsakilab.nyumc.org/datasets/TingleyD/>.

Code availability

The code used in this study is available from
<https://github.com/buzsakilab/buzcode>.

References

1. 1.

Kanoski, S. E. & Grill, H. J. Hippocampus contributions to food intake control: mnemonic, neuroanatomical, and endocrine mechanisms. *Biol.*

Psychiatry **81**, 748–756 (2017).

2. 2.

Lavenex, P. & Amaral, D. G. Hippocampal–neocortical interaction: a hierarchy of associativity. *Hippocampus* **10**, 420–430 (2000).

3. 3.

Seto, K., Saito, H., Otsuka, K. & Kawakami, M. Influence of electrical stimulation of the limbic structure on insulin level in rabbit's plasma. *Exp. Clin. Endocrinol.* **81**, 347–349 (1983).

4. 4.

Endroczi, E., Lissak, K., Bohus, B. & Kovacs, S. The inhibitory influence of archicortical structures on pituitary-adrenal function. *Acta Physiol. Acad. Sci. Hung.* **16**, 17–22 (1959).

5. 5.

Rubin, R. T., Mandell, A. J. & Crandall, P. H. Corticosteroid responses to limbic stimulation in man: localization of stimulus sites. *Science* **153**, 767–768 (1966).

6. 6.

McGaugh, J. L. & Gold, P. E. *Hormonal Modulation of Memory. Psychoendocrinology* (Academic, 1989).

7. 7.

Gold, P. E. & van Buskirk, R. Effects of posttrial hormone injections on memory processes. *Horm. Behav.* **7**, 509–517 (1976).

8. 8.

McEwen, B. S., Weiss, J. M. & Schwartz, L. S. Uptake of corticosterone by rat brain and its concentration by certain limbic

structures. *Brain Res.* **16**, 227–241 (1969).

9. 9.

Lathe, R. Hormones and the hippocampus. *J. Endocrinol.* **169**, 205–231 (2001).

10. 10.

Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).

11. 11.

Bohus, B. in *The Hippocampus: Volume 1: Structure and Development* (eds Isaacson, R. L. & Pribram, K. H.) 323–353 (Springer, 1975).

12. 12.

Jacobson, L. & Sapolsky, R. The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. *Endocr. Rev.* **12**, 118–134 (1991).

13. 13.

Hsu, T. M. et al. Hippocampus ghrelin signaling mediates appetite through lateral hypothalamic orexin pathways. *eLife* **4**, e11190 (2015).

14. 14.

Saphier, D. & Feldman, S. Effects of septal and hippocampal stimuli on paraventricular nucleus neurons. *Neuroscience* **20**, 749–755 (1987).

15. 15.

Prutskova, N. P. & Petrov, Yu. A. Electrophysiological investigation of the hippocampal projections to the neurosecretory cells of the

supraoptic nucleus of the rat hypothalamus. *Neurosci. Behav. Physiol.* **20**, 194–200 (1990).

16. 16.

Tasali, E., Leproult, R., Ehrmann, D. A. & Van Cauter, E. Slow-wave sleep and the risk of type 2 diabetes in humans. *Proc. Natl Acad. Sci. USA* **105**, 1044–1049 (2008).

17. 17.

Cappuccio, F. P., D’Elia, L., Strazzullo, P. & Miller, M. A. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* **33**, 414–420 (2010).

18. 18.

Spiegel, K., Leproult, R. & Van Cauter, E. Impact of sleep debt on metabolic and endocrine function. *Lancet* **354**, 1435–1439 (1999).

19. 19.

Amaral, D. G. & Lavenex, P. in *The Hippocampus Book* (eds. Anderson, P. et al.) 45–66 (Oxford Univ. Press, 2006).

20. 20.

Cenquizca, L. A. & Swanson, L. W. Analysis of direct hippocampal cortical field CA1 axonal projections to diencephalon in the rat. *J. Comp. Neurol.* **497**, 101–114 (2006).

21. 21.

Shimazu, T. Neuronal regulation of hepatic glucose metabolism in mammals. *Diabetes Metab. Rev.* **3**, 185–206 (1987).

22. 22.

Buzsáki, G. Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–1188 (2015).

23. 23.

Hansen, K. Oscillations in the blood sugar in fasting normal persons. *Acta Med. Scand.* **57**, 27–32 (1923).

24. 24.

Risold, P. Y. & Swanson, L. W. Connections of the rat lateral septal complex. *Brain Res. Brain Res. Rev.* **24**, 115–195 (1997).

25. 25.

Sosa, M., Joo, H. R. & Frank, L. M. Dorsal and ventral hippocampal sharp-wave ripples activate distinct nucleus accumbens networks. *Neuron* **105**, 725–741 (2020).

26. 26.

Tingley, D. & Buzsáki, G. Routing of hippocampal ripples to subcortical structures via the lateral septum. *Neuron* **105**, 138–149 (2020).

27. 27.

Niijima, A. Neural control of blood glucose level. *Jpn. J. Physiol.* **36**, 827–841 (1986).

28. 28.

Britton, S. W. Studies on the conditions of activity in endocrine glands: XVII. The nervous control of insulin secretion. *Am. J. Physiol.* **74**, 291–308 (1925).

29. 29.

Corkill, A. B., Marks, H. P. & White, W. E. Relation of the pituitary gland to the action of insulin and adrenaline. *J. Physiol.* **612**, 193–205 (1933).

30. 30.

Kuo, T., McQueen, A., Chen, T.-C. & Wang, J.-C. Regulation of glucose homeostasis by glucocorticoids. *Adv. Exp. Med. Biol.* **872**, 99–126 (2015).

31. 31.

Kim, S. H. & Park, M.-J. Effects of growth hormone on glucose metabolism and insulin resistance in human. *Ann. Pediatr. Endocrinol. Metab.* **22**, 145–152 (2017).

32. 32.

Güemes, A. & Georgiou, P. Review of the role of the nervous system in glucose homoeostasis and future perspectives towards the management of diabetes. *Bioelectron. Med.* **4**, 9 (2018).

33. 33.

Banks, W. A. & Kastin, A. J. Differential permeability of the blood-brain barrier to two pancreatic peptides: insulin and amylin. *Peptides* **19**, 883–889 (1998).

34. 34.

Schulingkamp, R. J., Pagano, T. C., Hung, D. & Raffa, R. B. Insulin receptors and insulin action in the brain: review and clinical implications. *Neurosci. Biobehav. Rev.* **24**, 855–872 (2000).

35. 35.

O’Malley, D., Shanley, L. J. & Harvey, J. Insulin inhibits rat hippocampal neurones via activation of ATP-sensitive K⁺ and large

conductance Ca^{2+} -activated K^+ channels. *Neuropharmacology* **44**, 855–863 (2003).

36. 36.

Harasta, A. E. et al. Septal glucagon-like peptide 1 receptor expression determines suppression of cocaine-induced behavior. *Neuropsychopharmacology* **40**, 1969–1978 (2015).

37. 37.

Somers, V. K., Dyken, M. E., Mark, A. L. & Abboud, F. M. Sympathetic-nerve activity during sleep in normal subjects. *N. Engl. J. Med.* **328**, 303–307 (1993).

38. 38.

Alford, F. P. et al. Temporal patterns of integrated plasma hormone levels during sleep and wakefulness. I. Thyroid-stimulating hormone, growth hormone and cortisol. *J. Clin. Endocrinol. Metab.* **37**, 841–847 (1973).

39. 39.

Logothetis, N. K. et al. Hippocampal–cortical interaction during periods of subcortical silence. *Nature* **491**, 547–553 (2012).

40. 40.

Swanson, R. A., Levenstein, D., McClain, K., Tingley, D. & Buzsáki, G. Variable specificity of memory trace reactivation during hippocampal sharp wave ripples. *Curr. Opin. Behav. Sci.* **32**, 126–135 (2020).

41. 41.

Csicsvari, J., Hirase, H., Czurkó, A., Mamiya, A. & Buzsáki, G. Fast network oscillations in the hippocampal CA1 region of the behaving rat. *J. Neurosci.* **19**, RC20 (1999).

42. 42.

Schomburg, E. W. et al. Theta phase segregation of input-specific gamma patterns in entorhinal–hippocampal networks. *Neuron* **84**, 470–485 (2014).

43. 43.

Watson, B. O., Levenstein, D., Greene, J. P., Gelinis, J. N. & Buzsáki, G. Network homeostasis and state dynamics of neocortical sleep. *Neuron* **90**, 839–852 (2016).

44. 44.

Levenstein, D., Watson, B. O., Rinzel, J. & Buzsáki, G. Sleep regulation of the distribution of cortical firing rates. *Curr. Opin. Neurobiol.* **44**, 34–42 (2017).

45. 45.

Chen, C. et al. Recent advances in electrochemical glucose biosensors: a review. *RSC Advances* **3**, 4473–4491 (2013).

46. 46.

Basu, A. et al. Time lag of glucose from intravascular to interstitial compartment in humans. *Diabetes* **62**, 4083–4087 (2013).

47. 47.

Kovatchev, B. P., Shields, D. & Breton, M. Graphical and numerical evaluation of continuous glucose sensing time lag. *Diabetes Technol. Ther.* **11**, 139–143 (2009).

48. 48.

Baker, D. A. & Gough, D. A. Dynamic delay and maximal dynamic error in continuous biosensors. *Anal. Chem.* **68**, 1292–1297 (1996).

49. 49.

Davey, R. J., Low, C., Jones, T. W. & Fournier, P. A. Contribution of an intrinsic lag of continuous glucose monitoring systems to differences in measured and actual glucose concentrations changing at variable rates in vitro. *J. Diabetes Sci. Technol.* **4**, 1393–1399 (2010).

50. 50.

Keenan, D. B., Mastrototaro, J. J., Voskanyan, G. & Steil, G. M. Delays in minimally invasive continuous glucose monitoring devices: a review of current technology. *J. Diabetes Sci. Technol.* **3**, 1207–1214 (2009).

51. 51.

Donoghue, T. et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020).

52. 52.

Tingley, D. & Buzsáki, G. Transformation of a spatial map across the hippocampal–lateral septal circuit. *Neuron* **98**, 1229–1242 (2018).

53. 53.

Stark, E. et al. Pyramidal cell–interneuron interactions underlie hippocampal ripple oscillations. *Neuron* **83**, 467–480 (2014).

54. 54.

Fernández-Ruiz, A. et al. Long-duration hippocampal sharp wave ripples improve memory. *Science* **364**, 1082–1086 (2019).

Acknowledgements

We thank M. Elmaleh, D. Nitz, A. Fernandez-Ruiz and A. Maurer for feedback on an early version of the manuscript. Funding was from NIH

MH122391, U19 NS104590 and U19NS107616.

Author information

Affiliations

1. Neuroscience Institute, New York University, New York, NY, USA
David Tingley, Ekin Kaya, Jordan Carpenter & György Buzsáki
2. Center for Neural Science, New York University, New York, NY, USA
Kathryn McClain & György Buzsáki
3. Department of Psychology, Bogazici University, Istanbul, Turkey
Ekin Kaya
4. Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Trondheim, Norway
Jordan Carpenter
5. Department of Neurology, New York University, New York, NY, USA
György Buzsáki
6. Langone Medical Center, New York University, New York, NY, USA
György Buzsáki

Contributions

D.T. and G.B. conceived the experiments. D.T., J.C., K.M. and E.K. performed the experiments presented. D.T. analysed data. D.T. and G.B. wrote the manuscript with contributions from other authors.

Corresponding authors

Correspondence to [David Tingley](#) or [György Buzsáki](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks H. Freyja Olafsdottir 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 Continuous glucose monitor quality control, circadian, mealtime and spectral properties.](#)

a, For a subset of experiments ($n = 43$ sessions in 8 rats) in which a continuous glucose monitor was implanted, the raw data are shown. Sorted by duration of patency. **b**, Average current readings (y axis) over time (x axis) after implantation. Note the slow decay over days, an indication of glucose oxidase enzyme breakdown on the electrode surface. **c**, Derivative (Δ current) of the data shown in **a**, showing faster time-scale fluctuations in glucose concentrations. **d**, Absolute value of the average Δ current (y axis) over time (x axis) after implantation. Note the relatively stable readings over days, and that the slow decay in absolute current readings is no longer present. **e**, Interstitial glucose concentrations are modulated over the circadian timescale. **f**, Interstitial glucose concentrations increase during the

postprandial period ($n = 36$ meals, collected from 4 rats). Meal size was one 3-g pellet of standard chow. **g**, Average power spectral density plot across rats ($n = 8$) of the Δ current signal. Red line is a linear best fit to the data. **h**, Subtraction of the linear fit in **c** from the observed power at each frequency band. Note the elevated power in the 15-min to 90-min range. Note also that the ultradian cycling of SPW-R and glucose levels are different (compare with Extended Data Fig. 3). **i**, Average auto-correlogram of glucose fluctuations across eight rats. Red dashed lines are the 99% confidence interval for each rat.

[Extended Data Fig. 2 SPW-R detection quality control.](#)

a, Auto-correlograms of all inter-ripple intervals. Each row is data from an individual rat. Note the peaks in the histogram from bursts of ripples, and the drop to zero, representing an approximately 20-ms refractory period between ripples. **b**, Heat map of the high-pass-filtered (>100 Hz) LFP for every SPW-R, sorted by peak amplitude of the event. **c**, Scatter plots of the z -scored ripple amplitude (x axis) and ripple frequency (y axis) for all ripples (black dots) and rats (rows).

[Extended Data Fig. 3 Circadian and ultradian rhythms in SPW-R rate.](#)

a, Histogram of SPW-R rate for data from eight rats and sessions ($n = 1,568.5$ recording hours from 8 rats; mean = 196 ± 84 h). **b**, Modulation of the SPW-R rate across the circadian timescale. Grey lines are individual rats, black line is the population average. **c**, Example auto-correlograms (± 12 h) of SPW-R rate for three rats. **d**, Average power spectral density plot ($n = 8$ rats) of ripple rate. Note two peaks at about 24 h and about 4 h. **e**, Cross-correlograms of detected SPW-Rs and CA1 population firing rates at two temporal scales. Left, 100-ms bins; right, 60-s median filter. At slow temporal scales (>250 ms), CA1 SPW-Rs are associated with an overall decrease in CA1 firing rates.

[Extended Data Fig. 4 SPW-R features and states that affect the correlation with glucose fluctuations.](#)

a, Histograms of duration, amplitude, frequency and inter-ripple intervals across all detected SPW-Rs. **b**, Average effect of SPW-R features on peri-SPW-R glucose fluctuations. The *y* axes are time (± 100 min) and the *x* axes are the percentile of each distribution (or time in milliseconds for ripple duration). White lines are the average effect at +10-min temporal offset. Red and blue asterisks indicate the column used for the plots in **c**. **c**, Red and blue traces highlight the effect on peri-SPW-R glucose fluctuations at two different positions in the distributions of ripple features in **b**. **d**, Cross-correlograms for four rats calculated on 4-h blocks of time in which the rat was fasted or fed. **e**, A sliding 2-h window of time was taken to calculate the cross-correlogram between SPW-Rs and glucose fluctuations before, during and after a 3-g meal was provided to fasted rats. **f**, The SPW-R–glucose fluctuation cross-correlogram was not different between fasting and fed states. **g**, Cross-correlograms for eight rats calculated using blocks of time in which the rat was in NREM sleep or awake. **h**, A sliding 2-h window of time was taken to calculate the cross-correlogram between SPW-Rs and glucose fluctuations, relative to brain state transitions. The negative correlation did not change in amplitude across either wake-to-NREM or NREM-to-wake state transitions. Bounds are one s.d. **i**, The SPW-R–glucose fluctuation cross-correlogram was not different between NREM and waking states. **j**, Average cross-correlograms between SPW-R rate and Δ current for 4-h windows centred at 06:00 h (blue), 12:00 h (red) and 20:00 (orange). **k**, Integrated cross-correlograms (5–20 min) over the 24-h light–dark cycle. **l**, Same line as in **j**, without s.d. bounds and zoomed in on the *y* axis. Note the cumulative effect size is only slightly larger during the dark period. Across rats, no hour of the day had a significantly different correlation than any other hour of the day (Student’s *t*-test; minimum $P = 0.53$). Note also that the circadian fluctuation of glucose level is not a mirror image of SPW-R rate (compare Extended Data Fig. [3b](#)).

Extended Data Fig. 5 Reverse correlation and oscillatory coupling of glucose fluctuations with SPW-R rate.

a, For each experiment peaks and troughs in glucose fluctuations were identified (red or black dotted lines; >1 s.d.). **b**, The average rate of SPW-Rs is shown when triggered on peaks in the glucose derivative (black line) and troughs in the glucose derivative (red line). Shaded area shows three

s.d. of a randomly circularly permuted null distribution (100 iterations). **c**, The average rate of stimulations is shown when triggered on peaks in the glucose derivative (black line) and troughs in the glucose derivative (red line). Shaded area, three s.d. of a randomly circularly permuted null distribution (100 iterations). **d**, Average power–power comodulogram at different frequencies of the glucose signal (*y* axis) and SPW-R rate signal (*x* axis) for 30 rats. Note that the patch of high correlations ($R > 0.2$) from 15–60 min is shifted upward, off of the diagonal. This indicates the SPW-R rate at slightly higher frequencies modulates glucose fluctuations at a slightly lower frequency, a hall-mark of phase-resetting mechanisms in biological systems. $P < 1^{-24}$ using a Student’s *t*-test with circularly shuffled data. **e**, Top, phase–amplitude coupling for each rat when using the phase angle of the glucose fluctuation signal (*x* axis) and SPW-R rate as the amplitude (colour; each rat is *z*-scored). Bottom, average phase–amplitude coupling across all rats. $P < 1^{-18}$ using the Rayleigh test for non-uniformity.

Extended Data Fig. 6 SPW-R rate is the best predictor of glucose concentration changes.

a, Leave-one-out prediction of glucose dynamics. Each heat map column represents the mean squared error difference between two models, one model with all predictors and a second with all predictors but one (leave-one-out analysis). The *y* axis is the temporal offset between predictors and glucose fluctuations (± 100 min in 5-min increments). The held-out predictor for each column is labelled along the *x* axis. **b**, Mean squared error difference for each model pair at +10-min offset. Each grey dot represents a single model pair with a random train and test split of the data (100 iterations). Note the averages for each model pair are equivalent to the +10 *y* axis of **a**. Box represents median, and 25th and 75th percentiles. **c**, Pairwise comparison of models using different predictors. For each pair of models using different predictors, the mean squared error values were subtracted. Each bin colour indicates the average mean squared error difference across 100 iterations with random train and test splits. **d**, For each pairwise predictor comparison, a two-sample *t*-test was performed across the 100 iterations with different train and test splits. The number and colour in each bin indicate the number of rats for which the mean squared error for the predictor column was significantly ($P < 0.05$) lower than the

mean squared error for the predictor row. For example, the upper leftmost value of 6 indicates that ripple rate was a better predictor than brain state in 6 of the 8 rats. **e**, Latent variables do not predict glucose fluctuations more strongly than SPW-R rate. The same pairwise approach was taken as in **c, d**, using a generalized linear model with two predictors to estimate glucose fluctuations. The predictors used were the principal components, or independent components, of all predictors listed in Fig. 3. SPW-R rate, by itself, was also included as an additional predictor. Columns are sorted by best prediction. For each pairwise comparison SPW-R rate allowed for better (or equivalent, for principal component 1) prediction. **f**, For each pair of models, the number of rats for which one model performed significantly better (two-way *t*-test $P < 0.05$) is shown. SPW-R rate and PC1 were similar in all rats, all other principal components and independent components performed worse for at least 4 out of 8 rats.

Extended Data Fig. 7 Stimulation protocol, statistics of optogenetic stimulation effect and control PPC experiments.

a, Autocorrelation of stimulation protocol showing no ultradian structure. **b**, Average stimulation rate was not different across the circadian light–dark cycle. **c**, Average P value of correlation between stimulation rate and Δ current at different temporal offsets (*x* axis) across all 8 rats. **d**, Number of rats (8 total) with $P < 0.05$ correlations between stimulation rate and Δ current at different temporal offsets (*x* axis). All tests were two-way *t*-tests. **e**, Optogenetic induction of population bursts in posterior parietal cortex does not induce decreases in glucose. Top, example histology. Bottom, example histology from the same experiment conducted in the PPC in 5 rats. **f**, Top, stimulation effect from CA1 experiments (as in Fig. 4). Bottom, stimulation effect from the same stimulation protocol in PPC. **g**, Red line is the average cross-correlogram between CA1 artificial ripples and glucose fluctuations. Blue line is the average cross-correlogram between PPC artificial ripples and glucose fluctuations. **h**, P values from a two-way *t*-test when comparing the cross-correlograms from the CA1 and PPC cohorts of rats. **i**, For reference, the cross-correlogram between artificial ripples and interstitial glucose fluctuations are shown (average from 8 rats). **j**, Average cross-correlogram between artificial ripples and the power spectrum slope of the LFP recorded in CA1. **k**, Average cross-correlogram between

artificial ripples and the high-pass-filtered coherence (a proxy for EMG⁴²). **l**, Average cross-correlogram between artificial ripples and movement as detected by an accelerometer on the head of the rat. **m**, Average cross-correlogram between artificial ripples and the ratio of wake to NREM sleep in each 5-min bin.

Extended Data Fig. 8 Nonlinear rectified response of lateral septum neurons to hippocampal synchrony.

a, Dorsal CA1 and CA3 projections converge within the lateral septum. Top, injection sites. Bottom, anterograde projections to the lateral septum. **b**, Left, black trace is one second of raw LFP recorded from CA1 pyramidal layer during running on a maze. Each row in the black raster plot indicates the action potentials from a single neuron. Grey dashed lines indicate the separation of theta cycles. The numbers below indicate the number of cells participating, and number of action potentials occurring, during each theta cycle. Right, three example SPW-Rs during sleep from the same session. In all example traces, neurons are sorted by the ordering of place fields on the maze (same sorting as left). **c**, Top, the proportion of theta cycles (black) and SPW-Rs (red) as a function of the hippocampal synchrony distribution. Dashed black and red vertical lines mark the position along the population synchrony axis at which 75% of all theta cycles (from left) or 75% of all SPW-Rs (from right) occurred. Bottom, the average z-scored firing rate across the lateral septum population (minimum threshold of 15 neurons) is shown as a function of the percentile of hippocampal synchrony distribution ($n = 38$ recordings from 5 rats). Bounds are one s.d.

Extended Data Fig. 9 Dorsal CA1 SPW-Rs more strongly correlate with glucose fluctuations than do ventral CA1 SPW-Rs.

a, Example histological verification of dorsal and ventral CA1 recording sites. Red triangles indicate the approximate location of the CA1 SPW-R detection channels for each region. Grey triangle indicates dorsal CA3. **b**, Example traces of raw LFP and bandpass-filtered LFP across both recording sites (top, ventral; bottom, dorsal). For each region, one shank

with uniformly distributed recording sites along the dorsal–ventral axis is shown. **c**, Example LFP from channels used for ripple detection. **d**, Average cross-correlogram between dorsal CA1 and ventral CA1 SPW-Rs at slow timescales. Inset, average cross-correlogram between dorsal CA1 and ventral CA1 SPW-Rs at fast timescales. **e**, Across the three rats with simultaneous dorsal and ventral hippocampal recordings, dorsal SPW-Rs were equivalently correlated with glucose fluctuations as the previous cohorts of rats with only dorsal CA1 recordings ($n = 30$). Ventral SPW-Rs had a significantly (two-way t -test) weaker correlation with peripheral glucose fluctuations. **f**, For each rat, ventral CA1 SPW-Rs were more weakly correlated with glucose dynamics and dorsal CA1 SPW-Rs.

[Extended Data Fig. 10 Histology from DREADD and CNO experiments.](#)

a, Top, histology from the cohort of rats with AAV1-hDlx-GiDREADD-dTomato-Fishell injected into lateral septum. Bottom, example cross-correlograms between SPW-R and glucose fluctuations on vehicle (blue) and CNO (black) days. **b**, Top, histology from the cohort of rats with AAV2-hSyn-hM4D(Gi)-mCherry injected into medial septum. Bottom, example cross-correlograms between SPW-R and glucose fluctuations on vehicle (blue) and CNO (black) days. **c**, Comparison of effects for individual rats across vehicle versus CNO days for each cohort. **d**, P values (two-way t -test) for each part of the cross-correlograms shown in **e** when comparing within hM4Di rats (red; CNO versus vehicle) or across cohorts of rats (blue; CNO–hM4Di vs CNO–control). **e**, Ten colleagues who were not involved with these experiments were provided histology from all 9 rats (6 lateral septum; 3 medial septum) and asked to rate the degree of viral transfection in each subregion of the septum. Each rat had three images taken along the anterior–posterior axis and the individuals were blinded as to which virus had been injected and the effects of CNO on the SPW-R–glucose correlation. For each rat and subregion, the average score from these 10 individuals is shown in the table. For each subregion, the correlation between these scores and the effect of CNO was calculated and is shown in bold below. Red values had a P values < 0.05 . **f**, Scatter plot of the SPW-R–glucose correlation after CNO injection (x axis) and the average expression score given by these rates (y axis) for the dorsal lateral

septum. g, Scatter plot of the SPW-R–glucose correlation after CNO injection (x axis) and the average expression score given by these rates (y axis) for the medial septum.

Supplementary information

Supplementary Information

This file contains Supplementary Table 1, the Supplementary Discussion and Supplementary References.

Reporting Summary

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Tingley, D., McClain, K., Kaya, E. *et al.* A metabolic function of the hippocampal sharp wave-ripple. *Nature* **597**, 82–86 (2021).
<https://doi.org/10.1038/s41586-021-03811-w>

- Received: 02 November 2020
- Accepted: 07 July 2021
- Published: 11 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03811-w>

Further reading

- [**A brain signal that coordinates thought with metabolism**](#)

- Manfred Hallschmid
- & Jan Born

Nature (2021)

[**A brain signal that coordinates thought with metabolism**](#)

- Manfred Hallschmid
- Jan Born

News & Views 11 Aug 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03811-w>

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

- Article
- [Published: 25 August 2021](#)

Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche

- [David Willnow^{1,2}](#),
- [Uwe Benary³](#),
- [Anca Margineanu](#) [ORCID: orcid.org/0000-0002-6634-9729³](#),
- [Maria Lillina Vignola](#) [ORCID: orcid.org/0000-0001-7121-7715¹](#),
- [Fabian Konrath](#) [ORCID: orcid.org/0000-0002-6843-6252³](#),
- [Igor M. Pongrac³](#),
- [Zahra Karimaddini^{4,5}](#),
- [Alessandra Vigilante¹](#),
- [Jana Wolf](#) [ORCID: orcid.org/0000-0003-3254-5868^{3,6}](#) &
- [Francesca M. Spagnoli](#) [ORCID: orcid.org/0000-0001-7094-8188¹](#)

[Nature](#) volume **597**, pages 87–91 (2021)

- 2405 Accesses
- 69 Altmetric
- [Metrics details](#)

Subjects

- [Developmental biology](#)
- [Stem-cell niche](#)

Abstract

Studies based on single cells have revealed vast cellular heterogeneity in stem cell and progenitor compartments, suggesting continuous differentiation trajectories with intermixing of cells at various states of lineage commitment and notable degrees of plasticity during organogenesis^{1,2,3,4,5}. The hepato-pancreato-biliary organ system relies on a small endoderm progenitor compartment that gives rise to a variety of different adult tissues, including the liver, pancreas, gall bladder and extra-hepatic bile ducts^{6,7}. Experimental manipulation of various developmental signals in the mouse embryo has underscored important cellular plasticity in this embryonic territory⁶. This is reflected in the existence of human genetic syndromes as well as congenital malformations featuring multi-organ phenotypes in liver, pancreas and gall bladder⁶. Nevertheless, the precise lineage hierarchy and succession of events leading to the segregation of an endoderm progenitor compartment into hepatic, biliary and pancreatic structures have not yet been established. Here we combine computational modelling approaches with genetic lineage tracing to accurately reconstruct the hepato-pancreato-biliary lineage tree. We show that a multipotent progenitor subpopulation persists in the pancreato-biliary organ rudiment, contributing cells not only to the pancreas and gall bladder but also to the liver. Moreover, using single-cell RNA sequencing and functional experiments we define a specialized niche that supports this subpopulation in a multipotent state for an extended time during development. Together these findings indicate sustained plasticity underlying hepato-pancreato-biliary development that might also explain the rapid expansion of the liver while attenuating pancreato-biliary growth.

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: Modelling of tissue dynamics in hepato-pancreatic organ rudiments.

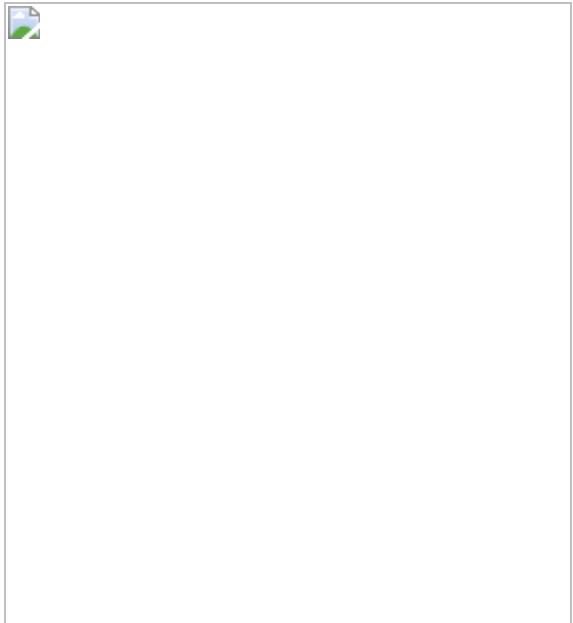


Fig. 2: Lineage tracing of hepato-pancreato-biliary progenitor populations.

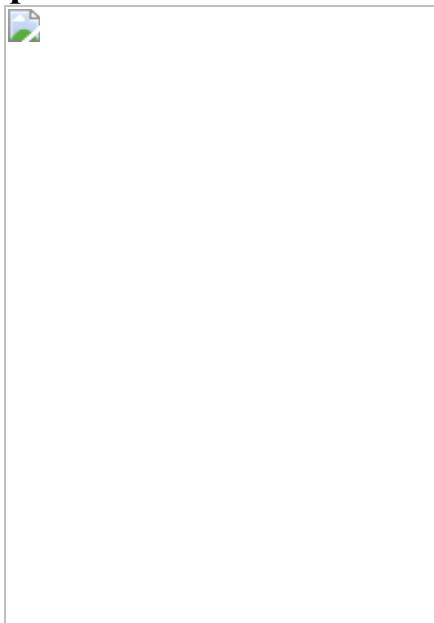


Fig. 3: scRNA-seq identifies a distinct progenitor subpopulation in the PB bud.

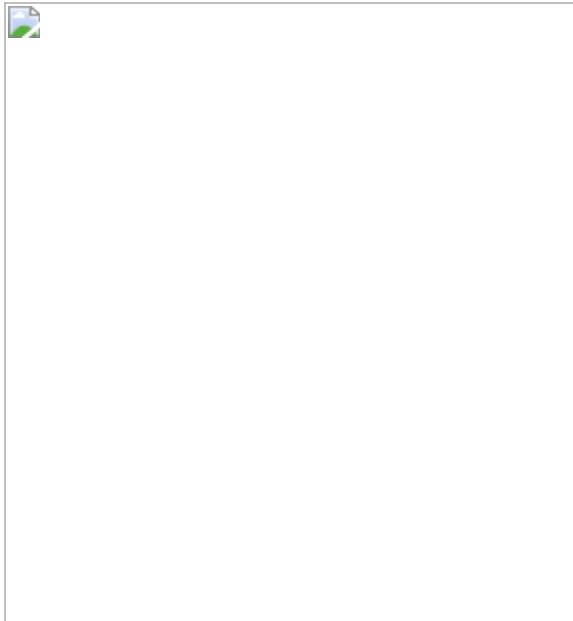
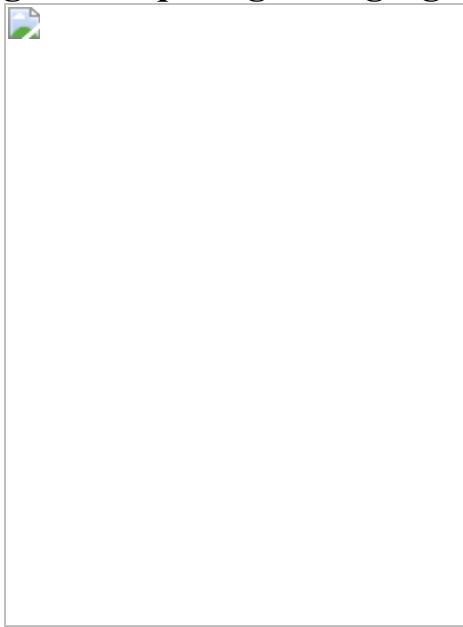


Fig. 4: Unique signalling signature defines the IMP population.



Data availability

The single-cell RNA-sequencing data in this study have been deposited in the Gene Expression Omnibus under accession code [GSE144103](#). [Source data](#) are provided with this paper.

Code availability

The single-cell RNA-sequencing data were analysed using publicly available R packages. Model D is available at
<https://www.ebi.ac.uk/biomodels/MODEL2105030001>.

References

1. 1.

Etzrodt, M., Endelev, M. & Schroeder, T. Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* **15**, 546–558 (2014).

2. 2.

Fischer, D. S. et al. Inferring population dynamics from single-cell RNA-sequencing time series data. *Nat. Biotechnol.* **37**, 461–468 (2019).

3. 3.

Haas, S., Trumpp, A. & Milsom, M. D. Causes and consequences of hematopoietic stem cell heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).

4. 4.

Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).

5. 5.

Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).

6. 6.

Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science* **322**, 1490–1494 (2008).

7. 7.

Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).

8. 8.

Slack, J. M. W. Origin of stem cells in organogenesis. *Science* **322**, 1498–1501 (2008).

9. 9.

Larsen, H. L. et al. Stochastic priming and spatial cues orchestrate heterogeneous clonal contribution to mouse pancreas organogenesis. *Nat. Commun.* **8**, 605 (2017).

10. 10.

Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).

11. 11.

Baron, C. S. & van Oudenaarden, A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* **20**, 753–765 (2019).

12. 12.

Pei, W. et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).

13. 13.

Rulands, S. & Simons, B. D. Tracing cellular dynamics in tissue development, maintenance and disease. *Curr. Opin. Cell Biol.* **43**, 38–45 (2016).

14. 14.

Prior, N. et al. Lgr5⁺ stem and progenitor cells reside at the apex of a heterogeneous embryonic hepatoblast pool. *Development* **146**, dev174557 (2019).

15. 15.

Tremblay, K. D. & Zaret, K. S. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev. Biol.* **280**, 87–99 (2005).

16. 16.

Li, L.-C. et al. Single-cell transcriptomic analyses reveal distinct dorsal/ventral pancreatic programs. *EMBO Rep.* **19**, e46148 (2018).

17. 17.

Spence, J. R. et al. Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev. Cell* **17**, 62–74 (2009).

18. 18.

Zhou, Q. et al. A multipotent progenitor domain guides pancreatic organogenesis. *Dev. Cell* **13**, 103–114 (2007).

19. 19.

Sosa-Pineda, B., Wigle, J. T. & Oliver, G. Hepatocyte migration during liver development requires Prox1. *Nat. Genet.* **25**, 254–255 (2000).

20. 20.

Raue, A. et al. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE* **8**, e74335 (2013).

21. 21.

Rodríguez-Seguel, E. et al. Mutually exclusive signaling signatures define the hepatic and pancreatic progenitor cell lineage divergence.

Genes Dev. **27**, 1932–1946 (2013).

22. 22.

Ober, E. A. & Lemaigre, F. P. Development of the liver: Insights into organ and tissue morphogenesis. *J. Hepatol.* **68**, 1049–1062 (2018).

23. 23.

Liu, J. et al. Non-parallel recombination limits Cre-LoxP-based reporters as precise indicators of conditional genetic manipulation. *Genesis* **51**, 436–442 (2013).

24. 24.

Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & McMahon, A. P. Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**, 1323–1326 (1998).

25. 25.

Lotto, J. et al. Single-cell transcriptomics reveals early emergence of liver parenchymal and non-parenchymal cell lineages. *Cell* **183**, 702–716.e14 (2020).

26. 26.

Hebrok, M. Hedgehog signaling in pancreas development. *Mech. Dev.* **120**, 45–57 (2003).

27. 27.

Bangs, F. & Anderson, K. V. Primary cilia and mammalian hedgehog signaling. *Cold Spring Harb. Perspect. Biol.* **9**, a028175 (2017).

28. 28.

Dessaud, E., McMahon, A. P. & Briscoe, J. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* **135**, 2489–2503 (2008).

29. 29.

Martínez-Frías, M. L. et al. Tracheoesophageal fistula, gastrointestinal abnormalities, hypospadias, and prenatal growth deficiency. *Am. J. Med. Genet.* **44**, 352–355 (1992).

30. 30.

Mitchell, J. et al. Neonatal diabetes, with hypoplastic pancreas, intestinal atresia and gall bladder hypoplasia: search for the aetiology of a new autosomal recessive syndrome. *Diabetologia* **47**, 2160–2167 (2004).

31. 31.

Heij, H. A. & Niessen, G. J. Annular pancreas associated with congenital absence of the gallbladder. *J. Pediatr. Surg.* **22**, 1033 (1987).

32. 32.

Park, E. J. et al. System for tamoxifen-inducible expression of cre-recombinase from the *Foxa2* locus in mice. *Dev. Dyn.* **237**, 447–453 (2008).

33. 33.

Hingorani, S. R. et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).

34. 34.

Perl, A.-K. T., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during

formation of the lung. *Proc. Natl Acad. Sci. USA* **99**, 10482–10487 (2002).

35. 35.

Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).

36. 36.

Abe, T. et al. Establishment of conditional reporter mouse lines at ROSA26 locus for live cell imaging. *Genesis* **49**, 579–590 (2011).

37. 37.

Gong, S. et al. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).

38. 38.

Means, A. L., Xu, Y., Zhao, A., Ray, K. C. & Gu, G. A CK19(CreERT) knockin mouse line allows for conditional DNA recombination in epithelial cells in multiple endodermal organs. *Genesis* **46**, 318–323 (2008).

39. 39.

Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).

40. 40.

Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

41. 41.

Gong, S., Yang, X. W., Li, C. & Heintz, N. Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6K γ origin of replication. *Genome Res.* **12**, 1992–1998 (2002).

42. 42.

Petzold, K. M. & Spagnoli, F. M. A system for ex vivo culturing of embryonic pancreas. *J. Vis. Exp.* **(66)**, e3979 (2012).
<https://doi.org/10.3791/3979>.

43. 43.

Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).

44. 44.

Dewitz, C. et al. Nuclear organization in the spinal cord depends on motor neuron lamination orchestrated by catenin and afadin function. *Cell Rep.* **22**, 1681–1694 (2018).

45. 45.

Husson, F., Josse, J. & Pagès, J. Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data? Applied Mathematics Department, 1–10 (2010).

46. 46.

Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**, 1–36 (2014).

47. 47.

Raue, A. et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics* **31**, 3558–3560 (2015).

48. 48.

Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).

49. 49.

Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).

50. 50.

McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).

51. 51.

Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

52. 52.

Amezquita, R. A. et al. Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).

53. 53.

Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

54. 54.

Escot, S., Willnow, D., Naumann, H., Di Francescantonio, S. & Spagnoli, F. M. Robo signalling controls pancreatic progenitor identity by regulating Tead transcription factors. *Nat. Commun.* **9**, 5082 (2018).

55. 55.

Palaria, A., Angelo, J. R., Guertin, T. M., Mager, J. & Tremblay, K. D. Patterning of the hepato-pancreatobiliary boundary by BMP reveals heterogeneity within the murine liver bud. *Hepatology* **68**, 274–288 (2018).

56. 56.

Sanders, E. J., Varedi, M. & French, A. S. Cell proliferation in the gastrulating chick embryo: a study using BrdU incorporation and PCNA localization. *Development* **118**, 389–399 (1993).

57. 57.

van den Berg, G. et al. A caudal proliferating growth center contributes to both poles of the forming heart tube. *Circ. Res.* **104**, 179–188 (2009).

58. 58.

Clevers, H. & Watt, F. M. Defining adult stem cells by function, not by phenotype. *Annu. Rev. Biochem.* **87**, 1015–1027 (2018).

Acknowledgements

We thank all the members of the Spagnoli laboratory for their useful comments and suggestions on the study; H. Naumann for technical help; G. Pazour for IFT57 antibody; A. Christ and T. Willnow for the GLI2 and SMO antibodies; the MDC Transgenic Unit for technical help in generating the *Prox1*-rtTA mouse strain; and C. Beisel of the Genomics Facility of D-BSSE for NGS RNA sequencing. Maintenance of the two-photon microscopy setup was supported by the staff of the Advanced Light Microscopy technology platform via funding from the MDC in the

Helmholtz Association. This research was supported by funds from the Helmholtz Association (FMS, JW), European Union's Horizon 2020 Research and Innovation Programme (Pan3DP Grant agreement no. 800981) (FMS, DW); D.W. was a recipient of a BIH (Tr. PhD) fellowship. U.B. was a recipient of an EMBO Short-Term Fellowship (7853).

Author information

Affiliations

1. Centre for Stem Cells and Regenerative Medicine, King's College London, London, UK

David Willnow, Maria Lillina Vignola, Alessandra Vigilante & Francesca M. Spagnoli

2. Berlin Institute of Health, Berlin, Germany

David Willnow

3. Max-Delbrueck-Center for Molecular Medicine, Berlin, Germany

Uwe Benary, Anca Margineanu, Fabian Konrath, Igor M. Pongrac & Jana Wolf

4. Department of Biosystems Science and Engineering (D-BSSE), ETH Zurich, Basel, Switzerland

Zahra Karimaddini

5. Swiss Institute of Bioinformatics, Basel, Switzerland

Zahra Karimaddini

6. Department of Mathematics and Computer Science, Free University, Berlin, Germany

Jana Wolf

Contributions

F.M.S. and D.W. conceived the study, designed the experiments and wrote the manuscript. D.W. performed all the experiments. I.M.P. generated the *Prox1*-rtTA transgenic strain. A.M. helped with the Confetti lineage-tracing experiments and two-photon image acquisition. U.B., F.K. and J.W. developed the mathematical modelling approach. M.L.V., A.V. and Z.K. performed scRNA-seq data analyses.

Corresponding author

Correspondence to [Francesca M. Spagnoli](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Dominic Grun, Meritxell Hutch 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 Expression of hepato-pancreato-biliary markers during early cell fate specification.](#)

a, Images of mouse embryos at the indicated somite stages (ss) taken using a Zeiss Stemi 2000 stereomicroscope. Scale bar, 1mm. **b**, Representative

confocal whole-mount immunofluorescence (IF) images of embryos stained for Prox1. In addition to the ventral foregut region (VFG), Prox1 is expressed in the developing heart tube (HT) from 6ss onwards. Scale bar, 100 μ m. **c–e**, Representative confocal IF images of the VFG in E8.5-E8.75 (8-13ss) embryos. VFG shown as whole-mount in frontal (**c**) or lateral (**d**) views or in transverse cryosections (**e**). Prox1 (red), Pdx1 (blue), and Sox17 (green) mark hepatic (Prox1 $^{+}$) and pancreato-biliary (Prox1 $^{+}$ /Pdx1 $^{+}$ /Sox17 $^{+}$) progenitors in the VFG (outlined by dotted line). Insets in **c, e** show higher magnifications of boxed regions in grey scale for Pdx1 channel. Right panels in **d** show Pdx1 channel in grey scale highlighting the onset of Pdx1 expression at the VFG lip. Hoechst dye was used as nuclear counterstain. Scale bars, 100 μ m. **f**, Scatter plot showing the ratio of LV *versus* PB cell counts ($n = 166$ embryos). Quantification of cell numbers in LV and PB buds was performed on IF stained cryosections (see Fig. 1b). Data are representative of 3 or more biologically independent experiments with similar results.

Source data

Extended Data Fig. 2 Dynamics of cell proliferation and apoptosis in hepatic, ventral pancreato-biliary, and dorsal pancreatic organ rudiments.

a–c, Representative IF images of cryosections of mouse embryos at indicated somite stages stained for pH3 (green) and indicated markers. In **a**, Ecad (red) marks epithelial cells of the endoderm and ectoderm at 0ss. In **b**, Ecad (blue)/Prox1 (red) mark the ventral foregut (VFG) at 6ss. In **c**, Prox1 (left panel) marks liver (LV) cells; Pdx1/Prox1 (left panel) or Pdx1/Sox17 (right panel) mark pancreato-biliary PB cells. In **c**, left and right panels show consecutive sections of the same embryos. VFG, LV, and PB are outlined by white dotted lines. Hoechst dye was used as nuclear counterstain. NF, neural fold. Scale bars, 100 μ m. **d**, Quantification of pH3 $^{+}$ cells in PB buds on cryosections ($n = 26$ embryos). Dot plot shows the numbers of pH3 $^{+}$ cells as % of the total cell count in the PB bud from E8.5 to E11.0 (9-41ss). Proliferating PB cells were counted on consecutive sections of the same embryos stained either for Prox1/Pdx1/pH3 or

Sox17/Pdx1/pH3. Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns, not significant. **e**, Quantification of pH3⁺ dorsal (DP) and ventral (VP) pancreatic cells on cryosections ($n = 28$ embryos). Dot plot shows the % fraction of pH3⁺ cells relative to the total cell number in each respective organ rudiment (% of total cell count) from E9.0 to E11.0 (15–44ss). VP cells within the PB bud were defined as described in Extended Data Fig. 5. Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns. **f**, Representative whole-mount IF images of embryos at E9.0–E9.75 (16–26 ss) stained for pH3 (green), Prox1 (red) and Pdx1 (blue). LV and PB buds are outlined by dashed white lines. Scale bar, 100 μ m. **g**, Measurement of LV and PB bud volumes at the indicated somite stages ($n = 129$ embryos). 3D organ volumes were reconstructed by measuring the surface area on individual optical sections of confocal z-series of whole-mount IFs (Fig. 1a). Average surface areas in each embryo were then multiplied by tissue thickness. **h**, Quantification of pH3⁺ cell numbers in whole-mount IF images of LV and PB buds ($n = 71$ embryos) normalized to organ bud volume (mm³). pH3 levels were higher in pancreato-biliary rudiments as compared to liver buds at early stages (E8.5; $P = 0.0016$), while no significant differences between the two organ rudiments were observed at later stages. Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns. **i**, Representative image of 18ss embryo stained for the indicated markers shows BrdU (green) labelled cells following a 4h-labelling period. Prox1 (red) marks LV, Prox1/Pdx1 (blue) marks PB progenitors. LV and PB buds are outlined by a dashed white line. Scale bar, 100 μ m. **j**, BrdU incorporation in LV and PB progenitors of E8.5 and E10.5 embryos following the indicated BrdU pulse-chasing periods. Dot plot showing the fraction of BrdU-labelled cells relative to the total cell number in each organ rudiment (% of total cell count). Each dot represents the mean from an individual embryo [E8.5: n(1h) = 4, n(2h) = 6, n(4h) = 8; E10.5: n(0.5h) = 3, n(4h) = 5]. Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns. **k**, Graph illustrating cell proliferation rates as shown by the % of BrdU⁺ cells in E8.5 and E10.5 LV and PB organ domains at different labelling periods (Extended Data Fig. 2j). The cell cycle length (t_c) was estimated based on BrdU incorporation using linear regression^{56,57}. LV (E8.5) slope = 3.8%/h; estimated cell cycle

length = 26.6h (range with 95% confidence interval: 44.7h-19h). PB (E8.5) slope = 5.8%/h; estimated cell cycle length = 17.3h (range with 95% confidence interval: 29.6h-12.2h). LV (E10.5) slope = 5.5%/h; estimated cell cycle length = 18.2h (range with 95% confidence interval: 21h-16.1h). PB (E10.5) slope = 4.8%/h; estimated cell cycle length = 21h (range with 95% confidence interval: 33h-15.4h). No statistically significant differences in average cell cycle length were found between LV and PB progenitors at E8.5 ($P = 0.14$) or E10.5 ($P = 0.36$). Two-tailed linear regression t-test. **I**, Fraction of pH3⁺/BrdU⁺ cells in LV and PB buds as % of the total BrdU⁺ cell population at the indicated pulse-chasing period [n(0.5h) = 3, n(4h) = 3]. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn's multiple comparisons test. ns. **m**, Fraction of BrdU⁺ early and late mitotic cells in LV and PB buds as % of the total early or late mitotic cell population at the indicated pulse-chasing period [n(0.5h) = 3, n(4h) = 3]. pH3 IF staining intensity was used to identify early (low, punctate pH3 signal) and late (high pH3 signal) mitotic cells. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn's multiple comparisons test. ns. **n**, Fraction of BrdU⁺ cells in LV, PB and DP buds of E9.5 embryos following the indicated pulse-chasing period [n(2h) = 2, n(4h) = 5]. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn's multiple comparisons test. ns. **o**, Fraction of BrdU⁺ cells in DP, VP and GB of E10.5 embryos following the indicated pulse-chasing period ($n = 4$). Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn's multiple comparisons test. ns. **p, q**, Immunostaining for apoptotic cell marker cleaved caspase 3 (cCas3) in embryos at the indicated somite stages. Representative IF images of cryosections of E8.0 (5ss) Tg(*Prox1*-EGFP) embryo (**p**) stained for cCas3 (blue), Prox1 (red) and GFP (green) to detect VFG cells (Prox1⁺, GFP⁺) and E9.5 (22ss) embryo (**q**) stained for cCas3 (blue), Prox1 (red) and Pdx1 (blue). Insets show higher magnifications of boxed regions in grey scale for the indicated channels. Rare apoptotic cells are found in LV and PB bud (arrowheads) as well as surrounding mesenchyme (arrow). Hoechst, nuclear counterstain. Scale bar, 100μm. **r**, Quantification of apoptotic (cCas3⁺) cells in VFG, LV, and PB buds on stained cryosections ($n = 19$ embryos) as shown in **p, q**. Dot plot shows the fraction of cCas3⁺ cells as % of total cell count in the indicated organ domains. No statistically significant differences in apoptosis were found between LV and PB buds at E8.5-E9.0 (12-20ss) or

E9.5-E10.0 (21-31ss). Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns. Data are representative of 3 or more biologically independent experiments with similar results.

[Source data](#)

Extended Data Fig. 3 Extended computational modelling analysis of ventral foregut development.

a, Models A and B fitted to the experimental data. Best fit of models A (black solid line) and B (red solid line) and the respective standard deviations are shown. Cell count data (black dots) collected between E7.5 and E11.5 (0-45ss) were used to estimate the parameter values of the models. **b**, Models C and D fitted to the experimental cell count data (black dots). Best fit of model C (black solid line) and model D (red solid line) and the respective standard deviations are shown. **c–g**, Simulations of distinct cell populations (indicated above each panel) in model C (black solid line) and model D (red solid line). **h, i**, Simulations of model D predict a flux of cells from the pancreato-biliary (PB) population to the hepatic (LV) population (**h**) throughout the time frame of the model, and a delayed increasing flux in the reverse direction (**i**). Until the 30ss, the flux of cells from the PB to the LV domain exceeds that of LV cells to the PB domain. Afterwards the direction of the positive net flux is reversed. **j**, Sum of BEP, BEPp, LV, and LVp cell counts (LV*) simulated with model D (red line) overlaid with the exponential function (black dashed line) that was fitted to the simulated cell counts to estimate the doubling time (11.8h). **k**, Sum of PB, PBp, VP, VPP, GB, and GBp cell counts (PB*) simulated with model D (red line) overlaid with the exponential function (black dashed line) that was fitted to the simulated cell counts to estimate the doubling time (14.1h). **l**, Sum of BEP, BEPp, LV, and LVp cell counts (LV*, dark red line) and sum of PB, PBp, VP, VPP, GB, and GBp cell counts (PB*, dark green line) simulated with model D using the nominal parameter values. Light colored lines represent simulations of LV* and PB* cell counts after reducing plasticity by setting parameter k_{17} and k_{18} to 1/4th of their nominal parameter values. The overlay shows that a reduction of plasticity reduces LV* cell count and increases PB* cell count. **m**, Ratio of the sum of BEP, BEPp, LV, and LVp cell counts (LV*) and the sum of PB, PBp, VP, VPP,

GB, and GBp cell counts (PB*) simulated with model D (red line) is overlaid with LV*/PB* ratios simulated with model D in which parameters k_{17} and k_{18} were both reduced to 1/2, 1/4th and 1/8th of their original values. The overlay shows that a reduction of parameters k_{17} and k_{18} (that is, plasticity) reduces LV*/PB* cell count ratio.

[Source data](#)

Extended Data Fig. 4 Overview of hepato-pancreato-biliary cell populations and their attributed marker gene expression.

a, Bar graph showing the number of data points that were collected for each somite stage and used for parameterizing the mathematical models. **b**, Schematic representation of mathematical model C overlaid with overlapping coloured boxes that indicate marker gene expression of various hepato-pancreato-biliary subpopulations (Prox1, red; Sox17, green; Pdx1, blue; pH3, grey). **c, d**, Eight new cell populations extend model D: LVs, LVsp, PBs, PBsp, GBs, GBsp, VPs, and VPsp that correspond to labelled cells (see Supplementary Note 1). Two different experimental setups were simulated: multicolour Prox1-rtTA;Confetti lineage tracing and Pdx1-CreERT lineage tracing experiments. The two experimental setups, Prox1-rtTA;Confetti lineage tracing (**c**) and Pdx1-CreERT lineage tracing experiments (**d**), differ in the sets of cell populations that are affected by the spike-in event. Abbreviation: s, spike-in.

Extended Data Fig. 5 Hepatic and pancreato-biliary organ domains contain cell subpopulations with distinct marker expression profiles.

a–f, Single-cell measurement of fluorescence intensity (FI) of Pdx1 and Sox17 in pancreato-biliary (PB) rudiments at E8.5-E11.5 (9-45ss). Pdx1 and Sox17 FI values for individual cells were plotted. Embryos of similar somite stages were grouped as indicated (n, number of embryos; 9-14ss: n = 27, 6608 cells; 15-20ss: n = 15, 4310 cells; 21-26ss: n = 16, 5125 cells; 27-32ss: n = 16, 6644 cells; 33-38ss: n = 16, 13002 cells; 39-45ss: n = 6, 10588 cells). From E8.5 to E9.5 (9ss-26ss), the majority of cells in the PB

organ rudiment co-expressed Sox17 and Pdx1, while at E10.0-E11.5 (27ss-45ss) they acquired either Sox17^{high} or Pdx1^{high} identity. FI was measured using ImageJ/Fiji. Values were corrected by linear normalization within each embryo. Black dashed lines indicate sub-division of the progenitor populations based on Pdx1 and Sox17 FI levels. **g**, Representative IF images of cryosections of the PB organ domain from E8.5 (11ss) to E11.0 (40ss) embryos. Pdx1 (green) and Sox17 (red) mark PB progenitor cells. After E9.75 (about 27ss), the PB bud segregates into gall bladder (GB; marked by Sox17) and ventral pancreatic bud (VP; marked by Pdx1). Hoechst dye was used as nuclear counterstain. Scale bar, 100µm. **h, i**, Scatter plots displaying the number of cells in different PB rudiments against developmental time ($n = 96$ embryos). FI data were used to categorize the subpopulations according to the relative expression levels of Pdx1 and Sox17. Pdx1^{high} subpopulation corresponds to VP progenitors (Pdx1-FI >50, Sox17-FI \le 50) (green dots); Sox17^{high} corresponds to GB progenitors (Pdx1-FI \le 50, Sox17-FI >50) (red dots); Pdx1/Sox17-double positive at low (Pdx1-FI \le 50, Sox17-FI \le 50) or high levels (Pdx1-FI >50, Sox17-FI >50) correspond to PB progenitors (yellow dots). The three subpopulations exhibit distinct propagation kinetics as shown by plotting absolute cell counts in the individual organ domains (**h**) or fraction of each subpopulation as percentage of the total cell number in the PB bud (**i**) against somite stage. While the relative fraction of PB progenitors decreased as compared to GB and VP progenitors after E9.75 [26-27ss; (**i**)], their total cell number remained constant throughout the analysed time period (**h**). **j, k**, Representative IF stainings for Prox1 (red), Pdx1 (blue), and albumin (green) on cryosections of embryos at the indicated somite stages. Between 15ss and 19ss (**j**), a subset of liver (LV) progenitor cells is positive for the hepatic marker albumin (arrowheads). At 36ss (**k**), the majority of LV cells co-express Prox1 and albumin, but a fraction of albumin⁻ cells is found in the LV (arrows). At no analysed time point, albumin⁺/Prox1⁻ cells were found in the LV. Right (**j**) or bottom (**k**) panels show higher magnifications of the boxed region as merged or single channels. Hoechst was used as nuclear counterstain. Scale bars, 100µm. **l**, Quantification of albumin⁺ cells in the LV bud on IF-stained cryosections ($n = 23$ embryos), shown as % of the Prox1⁺ LV cells. **m**, Comparison of LV cell counts measured on cryosections stained for Prox1 and albumin

($n = 20$ embryos) with LV cell counts predicted from model D [Fig. 1e; predicted LV cell count combines model predictions for the LV and LVp (Extended Data Fig. 3d) populations]. Expression measurements agree well with the model predictions at early somite stages (<30ss), although the LV cell counts based on albumin expression (red dots) or Prox1 expression (black circles) differ slightly at these early time points.

[Source data](#)

[**Extended Data Fig. 6 Extended lineage tracing analysis of hepato-pancreato-biliary progenitor populations.**](#)

a, Schematic representation of the transgenic constructs used for the *Prox1*-rtTA lineage tracing of mouse hepato-pancreato-biliary progenitors. The 2A peptide enabled co-expression of the rtTA and mCherry fluorescent reporter transgene. loxP sites are shown as purple triangles. **b, c**, Assessment of experimental parameters influencing *Prox1*-rtTA labelling efficiency. Bar plot in **b** shows the percentage (%) of embryos with induced label at increasing doxycycline doses [75 $\mu\text{g}/\text{g}$ body weight (bw), $n = 22$; 100 $\mu\text{g}/\text{g}$ bw, $n = 29$; 150 $\mu\text{g}/\text{g}$ bw, $n = 116$]. Graph in **c** shows the positive correlation between the number of genetically labelled cells in individual embryos and somite stage of labelling for the three indicated doxycycline doses. For all further analyses, pregnant females were injected with 150 $\mu\text{g}/\text{g}$ bw of doxycycline. **d**, Plots displaying the number of genetically labelled hepatic (LV) or pancreato-biliary (PB) progenitor cells, as percentage (%) of total labelled cells in both rudiments, plotted against somite stage in *Foxa2*-CreERT experiments (upper panel) or tracing period in *Prox1*-rtTA experiments (lower panel). Paired dots vertically aligned correspond to labelled cells (as %) measured in the LV (red dot) or PB (green dot) of the same embryo. Two-tailed linear regression t-test; $P = 0.006$ (*Foxa2*-CreERT; $n = 83$ embryos), $P = 0.061$ (*Prox1*-rtTA; $n = 81$ embryos). **e**, Schematic representation of the transgenic constructs used for the *Foxa2*-CreERT lineage tracing of mouse hepato-pancreato-biliary progenitors. loxP sites are shown as purple triangles. IRES, internal ribosome entry site; mER, murine oestrogen receptor. **f**, Schematic representation of experimental setup for clonal labelling of Tg(*Foxa*-CreERT;R26R-H2B-GFP) embryos. Pregnant females were intraperitoneally (IP) injected with a

single dose of tamoxifen (TAM; 12 μ g/g bw) at E8.5 and embryos collected at E9.0-E9.5 ($n = 15$ embryos). **g**, Representative whole-mount IF of Tg(*Foxa2*-CreERT;R26R-H2B-GFP) embryos at E9.0-E9.5 (16-21ss) following TAM injection at E8.5. Arrowheads indicate genetically labelled (GFP $^+$) cells in LV and PB buds. In clonal labelling experiments, 2-3 labelled cells were often found in close proximity to one another, suggesting they were descendants of a common labelled progenitor cell. Boxed region is shown at a higher magnification in the right panels; white dotted circles indicate two adjacent GFP $^+$ cells, one Pdx1 $^+$ and one Pdx1 $^-$, which possibly arose from a common progenitor. Scale bar, 100 μ m. **h**, Schematic representation of the R26R-Confetti transgene. Following recombination, genetically labelled cells either express nuclear GFP (nGFP), cytoplasmic RFP or YFP, or membrane-associated CFP (mCFP). **i**, Representative optical section of two-photon microscopy 3D scans of E11.5 Tg(*Prox1*-rtTA;R26R-Confetti) embryos. Genetically labelled cells expressing CFP, YFP, or RFP were detected in both ventral (VP) and dorsal pancreatic (DP) buds (circled by dashed white lines). Scale bar, 100 μ m. **j**, Representative images of VP and DP buds from E10.5 Tg(*Pdx1*-Cre;R26R-H2B-GFP) (top) or Tg(*Pdx1*-Cre;R26R-Confetti) (bottom) embryos. Top panels, IF staining for Pdx1 (blue) and Prox1 (red) mark the VP and DP buds; GFP $^+$ (green) labelled cells are in both VP and DP of Tg(*Pdx1*-Cre;R26R-H2B-GFP) embryos. Bottom panels, optical sections of two-photon microscopy 3D scans of Tg(*Pdx1*-Cre;R26R-Confetti) embryos show native reporter fluorescence of CFP, YFP, and RFP in genetically labelled cells. Scale bars, 100 μ m. **k**, Quantitative analysis of the contribution of the indicated fluorophores to the total number of labelled cells in the embryos shows characteristic induction efficiencies for each fluorophore that are comparable among different tissues. Mean \pm s.d. Brown-Forsythe and Welch ANOVA test with Dunnett's T3 multiple comparisons test. $P < 0.001$. **l**, In silico re-construction of genetically labelled cells in pancreatic buds of Tg(*Prox1*-rtTA;R26R-Confetti) embryos. Spot detection analysis in Imaris software was used to identify xyz-coordinates of individual genetically labelled cells and to reconstruct the labelled tissues. **m**, **n**, Clone size distribution established from Confetti lineage tracing experiments using the *Prox1*-rtTA (**m**) or *Pdx1*-Cre (**n**) Tg lines. Information on xyz-coordinates of labelled cells in DP and VP was used for clustering individual cells into clonal clusters based on their

geometric distance from one another⁴⁶. Significantly different distribution of clone sizes was measured between VP and DP, with most clusters in the VP being composed of 4-5 cells, while about 8 cells per cluster were found in the DP. (*Prox1*, $n = 38$ embryos; *Pdx1*, $n = 16$ embryos). Vertical dotted lines indicate mean values; two-tailed Mann–Whitney test. $P < 0.001$. **o**, Quantification of total number of cells (grey dots) and GFP-labelled cells (green dots) in DP and VP of E10.5 Tg(*Pdx1-Cre;R26R-H2B-GFP*) embryos ($n = 18$ embryos). Total cell count and GFP⁺ labelled cells are significantly higher in the DP as compared to the VP. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. $P < 0.001$. **p**, Quantification of the GFP⁺ cell population as percentage (%) of the total cell population in VP (green dots) or DP (blue dots) of Tg(*Pdx1-Cre;R26R-H2B-GFP*) embryos ($n = 18$ embryos). No statistically significant differences were found between DP and VP. Mean ± s.d. Two-tailed Mann–Whitney test. ns, not significant. **q**, Quantification of mitotic H2B-GFP⁺ cells in LV, DP, and PB organ rudiments on IF-stained cryosections of E11.5 Tg(*Prox1-rtTA;R26R-H2B-GFP*) ($n = 4$ embryos) and E10.5 Tg(*Pdx1-Cre;R26R-H2B-GFP*) embryos ($n = 4$ embryos). Dot plot shows the fraction of genetically labelled cells in mitosis relative to the total labelled cell number in the respective organ rudiments. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ns. **r**, Simulation of *Prox1-rtTA;Confetti* lineage tracing experiment (see Supplementary Note 1, Extended Data Fig. 4c). Fraction of labelled cells in all VP* population (ratio of the sum of VPs and VPsp cell counts versus the sum of VPs, VPsp, VP, and VPp cell counts) is shown for different combinations of label induction time points and recombination efficiencies. Grey surface shows the fraction of labelled VP* simulated using model D with nominal parameter values; white-red gradient surface shows the fraction of labelled VP* simulated using model D after reducing plasticity by setting parameter k_{17} and k_{18} to 1/8th of their nominal values. The fraction of labelled VP* is always higher in the model with reduced plasticity, supporting the hypothesis that the PB to LV plasticity results in clonal dispersal and reduced cluster sizes in the VP. **s**, Sum of VP and VPp cell counts obtained from simulations with model D using the nominal k_{17} and k_{18} parameter values (red line) or simulations with model D, in which both parameters k_{17} and k_{18} were reduced to 1/2 (black line), 1/4th (dark grey line), or 1/8th

(light grey line) of their nominal values. The overlay shows that a reduction of parameters k_{17} and k_{18} (that is, plasticity) results in increased VP and VPP cell counts.

Source data

Extended Data Fig. 7 Extended lineage tracing analysis of Pdx1⁺ pancreato-biliary progenitors.

a–c, Spatial representation of *Pdx1*-Cre lineage tracing experiments. IF images of cryosections from E10.5 (**a, b**) and E11.0 (**c**) Tg(*Pdx1*-Cre; *R26R*-H2B-GFP) embryos stained for Prox1, Pdx1, and GFP were digitalized to obtain xyz-coordinates for all GFP[−] (red) and GFP⁺ (green) cells in the liver (LV) and pancreato-biliary (PB; blue) buds⁴⁴. Data from individual embryos at E10.5 (**a, b**; $n = 8$) and E11.0 (**c**; $n = 10$) were combined and cells plotted according to their xy- (**b**) or yz-coordinates (**a, c**). **d**, Representative IF image of Tg(*Pdx1*-Cre; *R26R*-H2B-GFP) embryo at E11.0. GFP (green) identifies labelled cells descended from Pdx1⁺ progenitors. Prox1 (red) marks the LV and gall bladder (GB; outlined by a white dashed line), Prox1/Pdx1 (blue) marks the ventral pancreas (VP; outlined by a white dashed line). Insets show higher magnification of boxed regions as merged and single GFP and Pdx1 channels. Arrows indicate GFP⁺ LV cells at the border with the VP and GB. Hoechst dye was used as nuclear counterstain. Scale bar, 100μm. **e**, Representative IF images of newborn Tg(*Pdx1*-Cre; *R26R*-H2B-GFP) mouse liver. Glutamine synthetase (GS, red) marks hepatocytes near the central vein, whereas Krt19 (blue) marks cholangiocytes. Co-staining of GFP with GS (arrows) or Krt19 (arrowheads) identifies descendants of Pdx1⁺ PB cells capable of differentiating into hepatocytes and cholangiocytes. Scale bar, 100μm. **f**, Representative IF image of Tg(*Pdx1*-Cre; *R26R*-H2B-GFP) embryo at E9.5. Right panels show higher magnification of the boxed region. Arrowheads indicate genetically labelled (GFP⁺) cells in PB bud and surrounding LV bud positive for Krt19. Scale bar, 100μm. **g**, Quantification of GFP⁺ cells in LV of genetically labelled Tg(*Pdx1*-Cre; *R26R*-H2B-GFP) embryos (E9.0, $n = 7$; E9.5, $n = 10$; E10.5, $n = 7$; E11.0, $n = 7$) and newborn mice (P1, $n = 5$) shows a constant labelling index of 0.8% of the total liver cell

population. No statistically significant differences were detected between analysed time points. Mean \pm s.d. Two-tailed Kruskal–Wallis test. ns, not significant. **h**, Quantification of labelled PB progenitors in Tg(*Pdx1-Cre;R26R-Confetti*) embryos (E9.5, $n = 4$; E10.5, $n = 10$). Genetically labelled cell populations increase with developmental stage. Mean \pm s.d. Two-tailed Mann–Whitney test. $P = 0.04$. **i**, Quantification of labelled LV progenitors as % of total labelled cell population in the ventral foregut endoderm of Tg(*Pdx1-Cre;R26R-Confetti*) embryos (E9.5, $n = 4$; E10.5, $n = 10$). Mean \pm s.d. Two-tailed Mann–Whitney test. $P = 0.03$. **j**, Representative IF image of cryosections from E10.5 Tg(*Pdx1-Cre;R26R-Confetti*) stained for Prox1 (blue), RFP (red) and GFP (green). Anti-GFP antibody also detects membrane-bound CFP and cytoplasmic YFP. Dashed white line marks PB progenitor cells. Insets show higher magnifications of the boxed regions highlighting genetically labelled descendants of Pdx1⁺ PB cells expressing YFP (arrowheads) or RFP (arrow) in the LV bud. Scale bar, 100 μ m. **k–n**, Spatial representations of clonal *Pdx1-Cre* lineage tracing experiments in four individual E10.5 Tg(*Pdx1-Cre;R26R-Confetti*) embryos. IF data were digitalized to obtain xyz-coordinates for individual labelled LV (LV-CFP, LV-RFP, LV-YFP) and PB progenitors (PB-cells)⁴⁴. Cells were plotted according to their xy- (**k–m**) or yz-coordinates (**n**). Individual clones of the same colour are distributed along hepatic chords. Yellow arrows suggest trajectories of the labelled cell clusters arising from Pdx1⁺ PB progenitors and spreading along hepatic chords.

Source data

Extended Data Fig. 8 Inducible lineage tracing experiments of Pdx1⁺ pancreato-biliary progenitors.

a, Schematic representation of the *Pdx1-CreERT* transgene and experimental setup to obtain genetically labelled Tg(*Pdx1-CreERT;R26R-H2B-GFP*) or Tg(*Pdx1-CreERT;R26R-tdTomato*) embryos. Pregnant females were intraperitoneally (IP) injected with a single dose of tamoxifen [TAM; 150 μ g/g (*R26R-H2B-GFP*) or 25 μ g/g (*R26R-tdTomato*) body weight] at E8.5, E9.5, or E10.5. Labelled embryos were collected at E11.5. Based on the reported time-delay between TAM administration and

induction for Cre-ERT2^{23,24}, Pdx1⁺ cells may be marked within a 24 h period post-injection. **b**, Representative IF images of cryosections from E11.5 Tg(*Pdx1*-CreERT;R26R-H2B-GFP) embryos stained for Prox1 (red), Pdx1 (blue), and GFP (green). GFP⁺ labelled cells were found in ventral pancreas (VP), gall bladder (GB), extrahepatic bile duct (EHBD), and liver (LV) following TAM administration at E8.5 (left panels) and E9.5 (right panels). Insets show higher magnifications of the boxed regions. GB, VP, and EHBD are outlined by white dashed lines. The EHBD structure is a derivative of the pancreato-biliary (PB) bud and becomes visible beginning at E11.5 in the mouse¹⁷. Scale bars, 100μm. **c**, Quantification showing the percentage (%) of labelled cells in VP of genetically labelled Tg(*Pdx1*-Cre;R26R-H2B-GFP) (grey dots; data in Extended Data Fig. 6p; n = 18 embryos) or Tg(*Pdx1*-CreERT;R26R-H2B-GFP) embryos [green dots; n(TAM E8.5) = 22 embryos, n(TAM E9.5) = 24 embryos, n(TAM E10.5) = 13 embryos]. The fraction of labelled VP cells induced by the constitutive *Pdx1*-Cre Tg was significantly higher compared to the TAM-inducible *Pdx1*-CreERT Tg, which is in line with previous reports^{23,39}. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. P < 0.001. ns, not significant. **d**, Quantification of the fraction of labelled non-VP (that is, in GB, EHBD, and LV) cells as percentage (%) of total labelled cells in Tg(*Pdx1*-Cre;R26R-H2B-GFP) (grey dots; n = 18 embryos) or Tg(*Pdx1*-CreERT;R26R-H2B-GFP) embryos [green dots; n(TAM E8.5) = 21 embryos, n(TAM E9.5) = 24 embryos, n(TAM E10.5) = 13 embryos]. The fraction of labelled non-VP cells induced with the *Pdx1*-Cre Tg was significantly higher compared to the TAM-inducible *Pdx1*-CreERT one. TAM induction at E8.5 led to a significantly higher fraction of labelled non-VP cells compared to TAM at E10.5. Mean ± s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test. ** P = 0.005; *** P < 0.001. **e**, Bar chart showing the percentage (%) of embryos with labelled LV cells following constitutive *Pdx1*-Cre [(18/19), 94.7%] or TAM-inducible *Pdx1*-CreERT lineage tracing strategies [TAM E8.5: (3/22), 15%; TAM E9.5: (2/24), 8.3%; TAM E10.5: (0/13), 0%]. **f**, Quantification showing the percentage (%) of labelled cells in VP of genetically labelled Tg(*Pdx1*-Cre;R26R-tdTomato) (grey dots; n = 5 embryos) or Tg(*Pdx1*-CreERT;R26R-tdTomato) embryos (red dots; n(TAM E8.5) = 7 embryos). The fraction of labelled VP cells induced by the constitutive *Pdx1*-Cre Tg was significantly higher compared to the

TAM-inducible *Pdx1*-CreERT Tg, which is in line with previous reports^{23,39}. Mean ± s.d. Two-tailed Mann–Whitney test. $P = 0.003$. **g**, Quantification of the fraction of labelled non-VP (that is, in GB, EHBD, and LV) cells as percentage (%) of total labelled cells in Tg(*Pdx1*-Cre;R26R-tdTomato) (grey dots; $n = 5$ embryos) or Tg(*Pdx1*-CreERT;R26R-tdTomato) embryos [red dots; n(TAM E8.5) = 7 embryos]. The fraction of labelled non-VP cells induced with the *Pdx1*-Cre Tg was significantly higher compared to the TAM-inducible *Pdx1*-CreERT one. Mean ± s.d. Two-tailed Mann–Whitney test. $P = 0.003$. **h**, Bar chart showing the percentage (%) of embryos with labelled LV cells following constitutive *Pdx1*-Cre [(5/5), 100%] or TAM-inducible *Pdx1*-CreERT lineage tracing strategies [TAM E8.5: (3/7), 42.9%]. **i**, Simulation of *Pdx1*-CreERT lineage tracing experiments (see Supplementary Note 1, Extended Data Fig. 4d). Fraction of labelled cells in LV* population (ratio of the sum of LVs and LVsp cell counts vs. the sum of LVs, LVsp, LV, and LVp cell counts) is shown for different combinations of label induction time point and recombination efficiency. The maximal fraction of labelled LV* is 21%. The fraction of labelled cells increases with higher recombination efficiency, reflecting the difference in labelled LV cells observed *in vivo* between *Pdx1*-Cre and *Pdx1*-CreERT lineage traced embryos (**c–h**). For any chosen recombination efficiency, the fraction of labelled LV cells decreases at later induction time points, corroborating the results from *Pdx1*-CreERT lineage tracing experiments (**d, e**). Specifically, induction time points after 30ss yield a low % of labelled LV cells, independent of the recombination efficiency. **j**, Simulations of three induction time points that correspond to the experimental setup (**a**) are shown.

Source data

Extended Data Fig. 9 Distinct marker gene signatures define hepatic and pancreato-biliary subpopulations.

a–g, Violin plots of normalized log-expression values of selected cell-type-specific marker genes (**a–d**) and genes encoding components of Fgfr2 and Robo-Slit (**e**), canonical and non-canonical Wnt (**f**) and Hedgehog signalling (**g**) pathways in distinct progenitor subpopulations from the sc-RNA-seq data set generated in this study. Data from E8.5 bipotent

endoderm progenitors (BEP), E10.5 hepatic progenitors (LV), E10.5 intermediate progenitors (IMP), and E10.5 ventral pancreato-biliary progenitor cells (PB cells) are shown. **h–k**, Marker gene expression projected on t-SNE plots. Cells are projected into t-SNE space, as in **n**, but are coloured by the relative expression of indicated hepato-pancreato-biliary marker genes instead of cluster assignment. Colours span a gradient from red (high expression) to grey (low expression). IMP cells are circled by a black line. **l, m**, Representative IF images of E10.5 embryos stained for Sox17 (red), Pdx1 (blue), Sox9 [green; (**l**)], and Hnf1 β [(green); (**m**)]. Right panels show higher magnifications of the boxed regions as merge (**l**) or single channels (**m**). In **l**, arrowheads mark cells bordering duodenum (DUO), LV, and PB buds (boxed region) showing low levels of Sox9 (green) and Pdx1 (blue). In (**m**), Hnf1 β is abundant in Sox17 $^+$ cells of the PB bud (boxed region 1) but low at the boundary between DUO, LV, and PB (boxed region 2). Scale bars, 100 μ m. **n**, Distinct hepatoblast and hepato-mesenchymal gene signatures define subpopulations in the E10.5 liver bud. t-SNE plot visualization of sc-RNA-seq from BEP, E10.5 LV, PB, dorsal pancreatic (DP) and E14.5 pancreatic cells (PAN) (see Fig. 3a). LV cells were found in two clusters, referred to as LV-A (dashed black line) and LV-B (solid black line). **o**, Heatmap of average expression levels of selected genes in E10.5 LV-A and LV-B subpopulations. LV-A displayed markers for hepatoblast cell type, while LV-B for hepato-mesenchymal hybrid progenitors, recently reported by Lotto et al. ²⁵. Both subpopulations expressed *Prox1*. **p–s**, Marker gene expression projected on t-SNE plots. Cells are projected into t-SNE space, as in **n**, but are coloured by their relative expression of indicated hepato-mesenchymal hybrid progenitor markers. Colours span a gradient from red (high expression) to grey (low expression).

Source data

Extended Data Fig. 10 Sc-RNA-seq data set integration of liver and pancreato-biliary progenitors.

a, t-SNE plot visualization of the publicly available sc-RNA-seq data set from Li et al. ¹⁶ of dorsal pancreatic cells [DP; E9.5 (30 cells), E10.5 (84 cells)], ventral pancreato-biliary cells [PB; E9.5 (44 cells), E10.5 (210

cells)], and hepatic progenitors [LV; E10.5 (22 cells)]. Cells are coloured by tissue of origin (upper panel), or embryonic stage (lower panel). **b–g**, Marker gene expression projected on t-SNE plots. Cells are projected into t-SNE space, as in **a**, and coloured by the relative expression of the indicated hepato-pancreato-biliary marker genes. Colours span a gradient from red (high expression) to grey (low expression). Intermediate progenitor (IMP) cells, which show an intermediate hepato-pancreato-biliary gene signature and locate between LV and PB clusters, are circled by solid black line. **h**, t-SNE visualization of Seurat integration of Li et al. ¹⁶ (dots) data set and sc-RNA-seq data generated in this study (triangles) (all data sets merged and scaled together). Cells are coloured by cluster identity (left panel; cluster assignment based on similarities in gene expression profiles) or by embryonic stage (right panel). A subset of E10.5 PB cells, from both data sets, shows an intermediate hepato-pancreato-biliary gene signature and locate between LV and PB clusters (see outlined single cells).

Abbreviations: EP, endocrine progenitors; PAN, pancreas. **i–n**, Marker gene expression projected on t-SNE plots. Cells are projected into t-SNE space, as in **h**, but are coloured by the relative expression of the indicated hepato-pancreato-biliary marker genes. Colours span a gradient from red (high expression) to grey (low expression). Dotted circles indicate IMP cells. **o**, Dot plot showing the expression of a subset of hepato-pancreato-biliary marker genes across cell clusters identified in the integrated data sets. Dot size represents fraction of cells in each cluster expressing the marker, colour shows mean expression levels. Colours span a gradient from blue (high expression) to red (low expression). **p**, Biological process GO term analysis for genes enriched in IMP cells compared to LV and PB progenitor populations in the merged data sets. **q–s**, E8.5 bipotent endoderm progenitor (BEP), E10.5 LV, and PB cells of the two integrated data sets [excluding DP and E14.5 pancreatic (PAN) cells in the original analysis] were used to construct a pseudotime trajectory. Monocle 3⁵³ arranged the sequenced cells into a branched trajectory with BEP as starting state and LV or VP/EP as end states. Major cell trajectories are visualized with UMAP but coloured as per cell cluster (**q**), pseudotime (**r**), or displaying IMP cells (**s**). The lines correspond to the principal graph learned by Monocle 3.

[Source data](#)

Extended Data Fig. 11 Extended characterization of developmental potential and signalling signatures in hepato-pancreato-biliary subpopulations.

a, Schematic representation of transgenes used for lineage tracing of intermediate progenitor (IMP) cells. **b**, Representative IF of E10.0 (29ss) cryosections stained for pH3, Pdx1, Prox1 (left), and Krt19 (right). Left and right panels show IFs of consecutive sections of the same embryo. DUO, duodenum; LV, hepatic progenitors ($\text{Prox}1^+$); PB, pancreato-biliary progenitors ($\text{Prox}1^+/\text{Pdx}1^+$; left panel, white dashed line); IMP ($\text{Pdx}1^+/\text{Krt}19^+$; right panel, white dashed line). Scale bar, 100 μm . **c**, Quantification of proliferating (pH3^+) cells in E10.0 (29-31ss) dorsal pancreas (DP), LV, PB and IMP domains on IF-stained cryosections (**b**). Graph shows the fraction of pH3^+ cells as percentage of the total cell number in each organ rudiment [n(LV, PB, DP) = 4 embryo; n(IMP) = 6 embryos]. Grey lines connect samples in which pH3^+ cells were counted on consecutive sections from the same embryo. No statistically significant differences in the size of the proliferative cell fractions were detected. Even though the IMP cells do not proliferate more than the surrounding compartments, it is possible that when they exit from their *niche* they might be recruited into a putative transit amplifying cell compartment and undergo increased rounds of division, as shown for example in the skin stem cell compartment⁵⁸. Mean \pm s.d. Two-tailed Kruskal–Wallis test with Dunn’s multiple comparisons test, ns. **d–h**, Spatial representation of *Krt19*-CreERT lineage tracing experiments. IF images of cryosections from E11.5 Tg(*Krt19*-CreERT; *R26R*-H2B-GFP) embryos stained for Prox1, Pdx1, and GFP were digitalized to obtain xyz-coordinates for GFP $^-$ (red) and GFP $^+$ (orange) liver (LV) as well as GFP $^-$ (blue) and GFP $^+$ (green) PB cells⁴⁴. Each plot shows data from an individual Tg(*Krt19*-CreERT; *R26R*-H2B-GFP) embryo with labelled cells in LV and PB organ domains. Cells are plotted according to their xy- (upper panels) or xz-coordinates (lower panels). Close proximity of labelled cells in different organs, such as in (d, f, g) strongly suggests that cells descended from a common multipotent progenitor at E9.5. **i, j**, Representative cryosections of E9.5-E10.5 embryos stained for the indicated markers. At E9.5, Flrt2 and Flrt3 co-localize in the PB bud, but mark distinct subpopulations within the PB rudiment at later

stages. At E10.0-E10.5, Flrt2 is enriched in the Pdx1⁺ ventral pancreas (VP), while Flrt3 is enriched in the Pdx1⁻ gall bladder (GB). Krt19 is enriched in intermediate (IMP) subpopulation at the border of DUO, LV, and PB buds. Scale bars, 100μm. **k**, **l**, Representative IF images of the E10.5 gut boundary between LV, PB and DUO for the indicated markers. Shh (green) is expressed in the DUO (**k**). Krt19 staining (green) defines the border zone in **l**, encompassing IMP cells. Right panels show higher magnifications of the boxed regions as merge or single channels of Shh (**k**) and Krt19 (**l**) staining. Dotted line demarcates the border between IMP and DUO. Hoechst dye was used as nuclear counterstain. Scale bars, 100μm. **m**, Representative images of the E10.5 embryonic neural tube stained for cilia marker Arl13b and Shh pathway components Smo and Gli2. The ventral neural tube region (vnt) served as positive control for active Hedgehog signalling^{[27,28](#)}, showing high density of Smo localization at the primary cilia (Arl13b⁺) together with Gli2 (bottom panel), while it is absent in the dorsal neural tube region (dnt)^{[27,28](#)}. Insets show higher magnification of boxed regions 1 and 2, arrowheads indicate colocalization of Gli2 and Arl13b at the tip of a subset of primary cilia in the vnt. **n**, IF for Arl13b (left) and Smo (right) on E10.5 hepato-pancreato-biliary region. DUO, PB, and IMP are outlined by dashed white lines. Hoechst dye was used as nuclear counterstain. The density of Smo⁺ primary cilia is high in cells of the DUO and IMP but low in the PB bud. Scale bars, 10μm. **o**, IF for Arl13b (red, left) and Smo (red, right) shows absence of Hedgehog signalling (that is, no ciliary Smo signal) in the DP, as marked by E-cadherin (Ecad; green). Right panels show higher magnifications of the boxed regions as for the indicated channels in grey scale. Hoechst dye was used as nuclear counterstain. Scale bar, 100μm. **p**, IF staining shows abundant levels of Shh (green) in the lung, vnt, and notochord (nc), as previously published^{[27,28](#)}. Hoechst dye was used as nuclear counterstain. Scale bar, 100μm. **q-t**, t-SNE plots showing the expression of *Flrt2* (**q**), *Flrt3* (**r**), *Robo2* (**s**), and *Ift57* (**t**) in our sc-RNA-Seq data set. Colours span a gradient from red (high expression) to grey (low expression).

[Source data](#)

Extended Data Fig. 12 Establishment of ventral foregut explants as a model system to manipulate the Hedgehog signalling pathway ex vivo.

a, Schematic representation of the experimental workflow to collect and culture ventral foregut explants from E9.5 mouse embryos. White dashed lines indicate incision sites for microdissection of the posterior ventral foregut and anterior midgut (containing the hepatic and pancreato-biliary organ rudiments) for ex vivo culture. **b**, Representative 3D images of whole-mount ventral foregut explants IF stained for Prox1, Pdx1, and Krt19. Prox1 marks hepatic (LV; red), Prox1/Pdx1 pancreatic (VP; white), and Prox1/Krt19 intermediate progenitor (IMP; green) domains. Explants were treated with 2 μ M Smoothened agonist (SAG) or 5 μ M KAAD-cyclopamine for 24h or left untreated (Control). Top panels show raw data in 3D projection mode. In middle panels, Prox1 $^+$ tissues were highlighted using the manual surface creation tool in Imaris (Bitplane) to quantify the volume of distinct hepato-pancreato-biliary organ domains. Bottom panels show 3D renderings for distinct organ domains generated and quantified using the automatic surface creation tool with manual thresholding in Imaris (Bitplane). Scale bars, 100 μ m. **c**, Representative images of whole-mount IF of a ventral foregut explant for Prox1 (red), Pdx1 (white), and Krt19 (green) (left panel). After antibodies elution, the same sample was re-stained for Pdx1 (white), Sox17 (red), and Cdx2 (green) to mark the gall bladder (GB; Pdx1 $^-$ /Sox17 $^+$; white dotted line) and intestine (Cdx2 $^+$) (right panel). **d–f**, Quantification of the volume of distinct organ domains in ventral foregut explants following SAG treatment (Control, $n = 12$; 2 μ M SAG, $n = 7$). Plots show LV/GB (**d**), GB/IMP (**e**), or GB/VP (**f**) volume ratios. Mean \pm s.d. SAG led to a reduced GB/IMP volume ratio [(e); two-tailed Mann–Whitney test; $P = 0.01$]. **g**, Schematic representation of the hepato-pancreato-biliary region in the E10.5 mouse embryo. The PB bud contains an IMP domain (green) that borders the LV and the duodenum and receives different signalling cues (green, red, blue arrows) from surrounding epithelial and mesenchymal tissues. Cells within this domain exhibit an extended cell fate plasticity during organogenesis (E8.5-E11.5) and retain the unique ability to contribute (black arrows) to the pancreatic bud (blue), GB (orange), and LV (red). Markers of the different progenitor

domains and signalling pathways involved in hepato-pancreato-biliary development are listed on the left.

[Source data](#)

Supplementary information

Supplementary Information

This file contains Supplementary Text, Supplementary Note 1 (Description of the mathematical models), Supplementary Note 2 (Generation of the Tg(Prox1-rtTA) transgenic mouse line) and Supplementary References

Reporting Summary

Description of the cell populations considered in the models

Supplementary Table 1 . Table including the description of the cell populations considered in the mathematical models, their acronyms, and attributed phenotypic markers related to Fig. 1, Extended Data Fig. 3, Supplementary Note 1.

Experimental data and sums of cell populations in models A and B

Supplementary Table 2 . Table showing the experimental data and their corresponding sums of cell populations in models A and B.

Experimental data and sums of cell populations in models C and D

Supplementary Table 3 . Table showing the experimental data and their corresponding sums of cell populations in models C and D.

Rate parameters of the models

Supplementary Table 4 . Table listing the rate parameters of the models [1/h].

Initial conditions of the models [cells]

Supplementary Table 5 . Table listing the initial cell numbers for each population in the different models.

Initial conditions of the extended model D and redistribution of cell counts at the time point of label induction

Supplementary Table 6 . Table describing the initial cell numbers for each population in extended model D and the rates at which cells are redistributed to spike in cell populations at the time point of label induction.

List of antibodies

Supplementary Table 7 . Table including all primary and secondary antibodies used in the study.

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. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

[**Source Data Extended Data Fig. 7**](#)

[**Source Data Extended Data Fig. 8**](#)

[**Source Data Extended Data Fig. 9**](#)

[**Source Data Extended Data Fig. 10**](#)

[**Source Data Extended Data Fig. 11**](#)

[**Source Data Extended Data Fig. 12**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Willnow, D., Benary, U., Margineanu, A. *et al.* Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche. *Nature* **597**, 87–91 (2021). <https://doi.org/10.1038/s41586-021-03844-1>

- Received: 09 January 2020
- Accepted: 21 July 2021
- Published: 25 August 2021
- Issue Date: 02 September 2021

- DOI: <https://doi.org/10.1038/s41586-021-03844-1>
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03844-1>

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

- Article
- [Published: 25 August 2021](#)

APRIL limits atherosclerosis by binding to heparan sulfate proteoglycans

- [Dimitrios Tsiantoulas](#) [ORCID: orcid.org/0000-0002-7743-3192](#)¹,
- [Mahya Eslami](#)²,
- [Georg Obermayer](#)^{1,3},
- [Marc Clement](#)⁴,
- [Diede Smeets](#)¹,
- [Florian J. Mayer](#)¹,
- [Máté G. Kiss](#) [ORCID: orcid.org/0000-0002-9215-8328](#)^{1,3},
- [Lennart Enders](#) [ORCID: orcid.org/0000-0001-8341-3350](#)³,
- [Juliane Weißer](#)³,
- [Laura Göderle](#) [ORCID: orcid.org/0000-0003-1037-3137](#)^{1,3},
- [Jordi Lambert](#)⁴,
- [Florian Frommlet](#)⁵,
- [André Mueller](#)³,
- [Tim Hendrikx](#)¹,
- [Maria Ozsvár-Kozma](#)^{1,3},
- [Florentina Porsch](#) [ORCID: orcid.org/0000-0002-2633-6632](#)^{1,3},
- [Laure Willen](#)²,
- [Taras Afonyushkin](#)^{1,3},
- [Jane E. Murphy](#) [ORCID: orcid.org/0000-0003-2201-9469](#)⁴,
- [Per Fogelstrand](#)⁶,
- [Olivier Donzé](#)⁷,
- [Gerard Pasterkamp](#)⁸,
- [Matthias Hoke](#)⁹,

- [Stefan Kubicek](#) [ORCID: orcid.org/0000-0003-0855-8343³](#),
- [Helle F. Jørgensen](#) [ORCID: orcid.org/0000-0002-7909-2977⁴](#),
- [Nicolas Danchin^{10,11}](#),
- [Tabassome Simon^{12,13}](#),
- [Hubert Scharnagl¹⁴](#),
- [Winfried März^{14,15,16}](#),
- [Jan Borén](#) [ORCID: orcid.org/0000-0003-0786-8091⁶](#),
- [Henry Hess¹⁷](#),
- [Ziad Mallat](#) [ORCID: orcid.org/0000-0003-0443-7878^{4,18}](#) [na1](#),
- [Pascal Schneider](#) [ORCID: orcid.org/0000-0003-0677-9409²](#) [na1](#) &
- [Christoph J. Binder](#) [ORCID: orcid.org/0000-0001-8313-7050^{1,3}](#)

Nature volume **597**, pages 92–96 (2021)

- 3411 Accesses
- 71 Altmetric
- [Metrics details](#)

Subjects

- [Cardiovascular diseases](#)
- [Cytokines](#)

Abstract

Atherosclerotic cardiovascular disease causes heart attacks and strokes, which are the leading causes of mortality worldwide¹. The formation of atherosclerotic plaques is initiated when low-density lipoproteins bind to heparan-sulfate proteoglycans (HSPGs)² and become trapped in the subendothelial space of large and medium size arteries, which leads to chronic inflammation and remodelling of the artery wall². A proliferation-inducing ligand (APRIL) is a cytokine that binds to HSPGs³, but the physiology of this interaction is largely unknown. Here we show that genetic ablation or antibody-mediated depletion of APRIL aggravates

atherosclerosis in mice. Mechanistically, we demonstrate that APRIL confers atheroprotection by binding to heparan sulfate chains of heparan-sulfate proteoglycan 2 (HSPG2), which limits the retention of low-density lipoproteins, accumulation of macrophages and formation of necrotic cores. Indeed, antibody-mediated depletion of APRIL in mice expressing heparan sulfate-deficient HSPG2 had no effect on the development of atherosclerosis. Treatment with a specific anti-APRIL antibody that promotes the binding of APRIL to HSPGs reduced experimental atherosclerosis. Furthermore, the serum levels of a form of human APRIL protein that binds to HSPGs, which we termed non-canonical APRIL (nc-APRIL), are associated independently of traditional risk factors with long-term cardiovascular mortality in patients with atherosclerosis. Our data reveal properties of APRIL that have broad pathophysiological implications for vascular homeostasis.

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: APRIL deficiency promotes atherosclerosis.



Fig. 2: APRIL protects against atherosclerosis by binding to HSPG2.

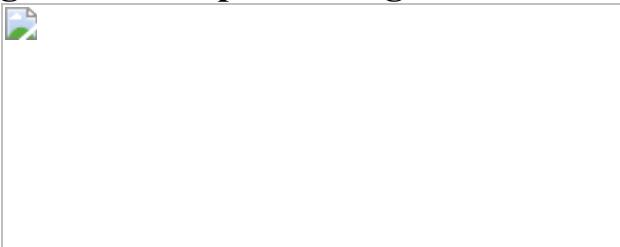


Fig. 3: Antibody targeting of APRIL at the BCMA/TACI binding site reduces atherosclerosis.

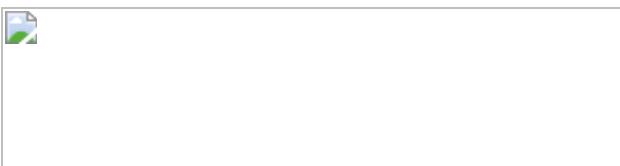
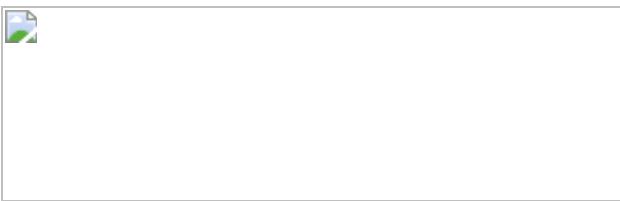


Fig. 4: Circulating nc-APRIL levels predict cardiovascular mortality in humans.



Data availability

The RNA sequencing datasets (from vascular smooth muscle cells) are available in the Gene Expression Omnibus with accession codes [GSE117963](#) and [GSE17858](#). All other relevant data are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

Code availability

For the clinical studies, the calculations were performed with SPSS (version 20.0, SPSS Inc.) for Windows, R version 3.6.0 (<https://www.R-project.org/>), and survival analysis was performed using the R packages survival (<https://CRAN.R-project.org/package=survival>) and survminer (<https://CRAN.R-project.org/package=survminer>). For analysis of mass spectrometry data, acquired raw data files were processed using Proteome Discoverer 2.4.1.15 SP1 for DDA experimental data or Skyline version 20.1.0.155 for PRM experimental data or using Mascot version 2.3.02 (Matrix Science, London, UK) and Phenyx (GeneBio, Geneva, Switzerland) as search engines. RNA-seq data were quality controlled using FastQC v0.11.3 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed using the Trim Galore v0.4.1 wrapper (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were aligned to the GRCm38 mouse reference genome using Tophat v2.0.12. Reads with a minimum map quality of 20 were imported into Seqmonk 1.45.4 (<http://www.bioinformatics.babraham.ac.uk/projects/seqmonk>).

References

1. 1.
Libby, P. The changing landscape of atherosclerosis. *Nature* **592**, 524–533 (2021).
2. 2.

Gisterå, A. & Hansson, G. K. The immunology of atherosclerosis. *Nat. Rev. Nephrol.* **13**, 368–380 (2017).

3. 3.

Ingold, K. et al. Identification of proteoglycans as the APRIL-specific binding partners. *J. Exp. Med.* **201**, 1375–1383 (2005).

4. 4.

Vincent, F. B., Morand, E. F., Schneider, P. & Mackay, F. The BAFF/APRIL system in SLE pathogenesis. *Nat. Rev. Rheumatol.* **10**, 365–373 (2014).

5. 5.

Castigli, E. et al. Impaired IgA class switching in APRIL-deficient mice. *Proc. Natl Acad. Sci. USA* **101**, 3903–3908 (2004).

6. 6.

Huard, B. et al. APRIL secreted by neutrophils binds to heparan sulfate proteoglycans to create plasma cell niches in human mucosa. *J. Clin. Invest.* **118**, 2887–2895 (2008).

7. 7.

McCarron, M. J., Park, P. W. & Fooksman, D. R. CD138 mediates selection of mature plasma cells by regulating their survival. *Blood* **129**, 2749–2759 (2017).

8. 8.

Hymowitz, S. G. et al. Structures of APRIL-receptor complexes: like BCMA, TACI employs only a single cysteine-rich domain for high affinity ligand binding. *J. Biol. Chem.* **280**, 7218–7227 (2005).

9. 9.

Sandberg, W. J. et al. The tumour necrosis factor superfamily ligand APRIL (TNFSF13) is released upon platelet activation and expressed in atherosclerosis. *Thromb. Haemost.* **102**, 704–710 (2009).

10. 10.

Lord, M. S. et al. The multifaceted roles of perlecan in fibrosis. *Matrix Biol.* **68-69**, 150–166 (2018).

11. 11.

Tran-Lundmark, K. et al. Heparan sulfate in perlecan promotes mouse atherosclerosis: roles in lipid permeability, lipid retention, and smooth muscle cell proliferation. *Circ. Res.* **103**, 43–52 (2008).

12. 12.

Sarrazin, S., Lamanna, W. C. & Esko, J. D. Heparan sulfate proteoglycans. *Cold Spring Harb. Perspect. Biol.* **3**, a004952 (2011).

13. 13.

Parish, C. R. The role of heparan sulphate in inflammation. *Nat. Rev. Immunol.* **6**, 633–643 (2006).

14. 14.

Bernelot Moens, S. J. et al. Impact of the B cell growth factor APRIL on the qualitative and immunological characteristics of atherosclerotic plaques. *PLoS One* **11**, e0164690 (2016).

15. 15.

Haselmayer, P., Vigolo, M., Nys, J., Schneider, P. & Hess, H. A mouse model of systemic lupus erythematosus responds better to soluble TACI than to soluble BAFFR, correlating with depletion of plasma cells. *Eur. J. Immunol.* **47**, 1075–1085 (2017).

16. 16.

Tsiantoulas, D. et al. B cell-activating factor neutralization aggravates atherosclerosis. *Circulation* **138**, 2263–2273 (2018).

17. 17.

Tsiantoulas, D. et al. Increased plasma IgE accelerate atherosclerosis in secreted IgM deficiency. *Circ. Res.* **120**, 78–84 (2017).

18. 18.

Dishman, A. F. et al. Evolution of fold switching in a metamorphic protein. *Science* **371**, 86–90 (2021).

19. 19.

Schillinger, M. et al. Inflammation and Carotid Artery—Risk for Atherosclerosis Study (ICARAS). *Circulation* **111**, 2203–2209 (2005).

20. 20.

Winkelmann, B. R. et al. Rationale and design of the LURIC study—a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics* **2** (Suppl 1), S1–S73 (2001).

21. 21.

Puymirat, E. et al. Acute myocardial infarction: changes in patient characteristics, management, and 6-month outcomes over a period of 20 years in the FAST-MI program (French Registry of Acute ST-Elevation or Non-ST-Elevation Myocardial Infarction) 1995 to 2015. *Circulation* **136**, 1908–1919 (2017).

22. 22.

Kowalczyk-Quintas, C. et al. Generation and characterization of function-blocking anti-ectodysplasin A (EDA) monoclonal antibodies that induce ectodermal dysplasia. *J. Biol. Chem.* **289**, 4273–4285 (2014).

23. 23.

Chou, M. Y. et al. Oxidation-specific epitopes are dominant targets of innate natural antibodies in mice and humans. *J. Clin. Invest.* **119**, 1335–1349 (2009).

24. 24.

Kijani, S., Vázquez, A. M., Levin, M., Borén, J. & Fogelstrand, P. Intimal hyperplasia induced by vascular intervention causes lipoprotein retention and accelerated atherosclerosis. *Physiol. Rep.* **5**, e13334 (2017).

25. 25.

Tom, R., Bisson, L. & Durocher, Y. Transfection of HEK293-EBNA1 cells in suspension with linear PEI for production of recombinant proteins. *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.prot4977> (2008).

26. 26.

Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

27. 27.

Bossen, C. et al. TACI, unlike BAFF-R, is solely activated by oligomeric BAFF and APRIL to support survival of activated B cells and plasmablasts. *Blood* **111**, 1004–1012 (2008).

28. 28.

Kowalczyk-Quintas, C. et al. Inhibition of membrane-bound BAFF by the anti-BAFF antibody Belimumab. *Front. Immunol.* **9**, 2698 (2018).

29. 29.

Schneider, P., Willen, L. & Smulski, C. R. Tools and techniques to study ligand-receptor interactions and receptor activation by TNF superfamily members. *Methods Enzymol.* **545**, 103–125 (2014).

30. 30.

Manza, L. L., Stamer, S. L., Ham, A.-J. L., Codreanu, S. G. & Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742–1745 (2005).

31. 31.

Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

32. 32.

Rappaport, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

33. 33.

Olsen, J. V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).

34. 34.

Hellings, W. E., Moll, F. L., de Kleijn, D. P. & Pasterkamp, G. 10-years experience with the Athero-Express study. *Cardiovasc. Diagn. Ther.* **2**, 63–73 (2012).

35. 35.

Mayer, F. J. et al. Combined effects of inflammatory status and carotid atherosclerosis: a 12-year follow-up study. *Stroke* **47**, 2952–2958

(2016).

36. 36.

Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

Acknowledgements

We thank A. Fabry and T. Wenko for help with the *in vivo* experimental studies, and C. Friedl for help with confocal microscopy. This work was supported by grants from the Austrian Science Fund (SFB F54), the European Union (FP7 VIA) and the Leducq Foundation (TNE-20CVD03) to C.J.B., and by grants from the European Research Area Network on Cardiovascular Diseases (I4647) and the British Heart Foundation (RCAG/917) to D.T. D.T. is also supported by the Austrian Science Fund (I4963). P.S. is supported by the Swiss National Science Foundation (31003A_176256, 310030E_197000). Z.M. is supported by the British Heart Foundation (RCAM/104, RCAM-659, RRCAM.163), the British Heart Foundation Center for Research Excellence (RE/18/1/34212), the NIHR Cambridge Biomedical Research Centre (RG85315), the European Union (FP7 VIA; RCAG/430) and the European Research Council (ERC). T.H. is supported by a NWO Veni grant (91619012).

Author information

Author notes

1. These authors contributed equally: Ziad Mallat, Pascal Schneider

Affiliations

1. Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

Dimitrios Tsiantoulas, Georg Obermayer, Diede Smeets, Florian J. Mayer, Máté G. Kiss, Laura Göderle, Tim Hendrikx, Maria Ozsvár-Kozma, Florentina Porsch, Taras Afonyushkin & Christoph J. Binder

2. Department of Biochemistry, University of Lausanne, Epalinges, Switzerland

Mahya Eslami, Laure Willen & Pascal Schneider

3. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

Georg Obermayer, Máté G. Kiss, Lennart Enders, Juliane Weißer, Laura Göderle, André Mueller, Maria Ozsvár-Kozma, Florentina Porsch, Taras Afonyushkin, Stefan Kubicek & Christoph J. Binder

4. Division of Cardiovascular Medicine, Department of Medicine, University of Cambridge, Cambridge, UK

Marc Clement, Jordi Lambert, Jane E. Murphy, Helle F. Jørgensen & Ziad Mallat

5. Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Florian Frommlet

6. Institute of Medicine, University of Gothenburg, Göteborg, Sweden

Per Fogelstrand & Jan Borén

7. Adipogen Life Sciences, Epalinges, Switzerland

Olivier Donzé

8. University Medical Center Utrecht, Utrecht, the Netherlands

Gerard Pasterkamp

9. Department of Internal Medicine II, Medical University of Vienna,
Vienna, Austria

Matthias Hoke

10. Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Européen
Georges Pompidou, Department of Cardiology, Paris, France

Nicolas Danchin

11. University School of Medicine, Université de Paris, Paris, France

Nicolas Danchin

12. Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Saint
Antoine, Department of Clinical Pharmacology and Clinical Research
Platform of East of Paris (URCEST-CRB-CRC), Paris, France

Tabassome Simon

13. Department of Pharmacology, Sorbonne-Université (UPMC-Paris 06),
Paris, France

Tabassome Simon

14. Clinical Institute of Medical and Chemical Laboratory Diagnostics,
Medical University of Graz, Graz, Austria

Hubert Scharnagl & Winfried März

15. Medical Clinic V, Medical Faculty Mannheim, University of
Heidelberg, Mannheim, Germany

Winfried März

16. SYNLAB Academy, Synlab Holding Deutschland GmbH, Augsburg,
Germany

Winfried März

17. Translational Innovation Platform Immunology, Merck KGaA,
Darmstadt, Germany

Henry Hess

18. Université de Paris and INSERM U970, Paris Cardiovascular Research Center, Paris, France

Ziad Mallat

Contributions

D.T. conceived and designed the study, performed most of the experiments, analysed and interpreted data, and wrote the manuscript. M.E. and P.S. generated materials, performed experiments to characterize nc-APRIL and interpreted data. G.O., L.E., S.K., L.W., T.A., T.H., M.G.K., M.O.-K., L.G., F.P., J.E.M. and P.F. aided in mouse studies and provided technical assistance with the experiments. M.C. aided in immunofluorescence analyses. D.S. aided in mouse studies. J.L. and H.F.J. provided the RNA-seq data for mouse VSMCs. A.M. and J.W. performed the mass-spectrometry analysis and analysed the data. F.J.M. and F.F. performed statistical analysis of the ICARAS and LURIC clinical data, respectively. O.D. and J.B. provided reagents and technical expertise with the experiments, and critically revised the manuscript. M.H. was involved in the analysis of human samples. T.S and N.D. provided the samples from the FAST-MI clinical study and performed the statistical analysis of the data. H.S., W.M. and Z.M. provided the samples from the LURIC clinical study and measured nc-APRIL levels. G.P. provided materials. H.H. provided materials and critically revised the manuscript. Z.M. and P.S. contributed to study design, interpreted data and critically revised the manuscript. C.J.B. designed the study, interpreted data and wrote the manuscript.

Corresponding authors

Correspondence to [Dimitrios Tsiantoulas](#) or [Christoph J. Binder](#).

Ethics declarations

Competing interests

D.T., C.J.B., P.S. and M.E. are named inventors on a patent application (EP20217536.0; pending) to exploit c-APRIL and nc-APRIL for diagnostic and therapeutic purposes in cardiovascular disease that has been filed by the Medical University of Vienna (Austria) and CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences (Austria). O.D. is an employee of Adipogen Life Sciences, which provided some reagents used in this study.

Additional information

Peer review information *Nature* thanks Peter Libby, William Sessa 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 APRIL deficiency does not alter the plaque smooth muscle cell and collagen content, or the numbers of circulating monocytes, B or T lymphocytes in *Ldlr*^{-/-} mice.

Ldlr^{-/-}*Tnfsf13*^{+/+} or *Ldlr*^{-/-}*Tnfsf13*^{-/-} mice were fed an atherogenic diet for 10 weeks. **a**, Whole body weight, plasma triglyceride levels and en face lesion size ($n = 10$ *Ldlr*^{-/-}*Tnfsf13*^{+/+} mice, $n = 12$ *Ldlr*^{-/-}*Tnfsf13*^{-/-} mice). **b-d**, Representative photomicrographs of DAPI- (**b**), α -SMA- (**c**) and Sirius Red-stained lesions (**d**) in the aortic origin (left) and dot plots (right) showing the averaged acellular (**b**; $n = 9$ *Ldlr*^{-/-}*Tnfsf13*^{+/+} mice, $n = 12$ *Ldlr*^{-/-}*Tnfsf13*^{-/-} mice, $P = 0.036$), α -SMA-positive (**c**, $n = 9$ *Ldlr*^{-/-}*Tnfsf13*^{+/+} mice, $n = 10$ *Ldlr*^{-/-}*Tnfsf13*^{-/-} mice) and collagen-positive area normalized to total lesion size (**d**; $n = 10$ *Ldlr*^{-/-}*Tnfsf13*^{+/+} mice,

$n = 12$ $Ldlr^{-/-} Tnfsf13^{-/-}$ mice). **e–g**, Representative flow cytometry dot plots (top) and quantification (bottom) of the absolute numbers of splenic CD3 $^{+}$, CD4 $^{+}$ (defined as CD3 $^{+}$ CD4 $^{+}$ CD8 $^{-}$) and CD8 $^{+}$ T cells (defined as CD3 $^{+}$ CD8 $^{+}$ CD4 $^{-}$) (**e**), frequencies of peritoneal B-1a (defined as B220 $^{\text{low}}$ CD11b $^{\text{int}}$ CD5 $^{+}$), B-1b (defined as B220 $^{\text{low}}$ CD11b $^{\text{int}}$ CD5 $^{-}$) and CD23 $^{+}$ B-2 (defined as B220 $^{\text{high}}$ CD11b $^{-}$ CD5 $^{-}$ CD23 $^{+}$) cells (**f**) and the frequencies of circulating Ly6C $^{\text{high}}$, Ly6C $^{\text{int}}$ and Ly6C $^{\text{low}}$ monocytes (**g**) in peripheral blood. All results show mean (two-tailed unpaired Student's *t*-test). Scale bars, 200 μm .

Source data

Extended Data Fig. 2 BCMA is dispensable for atherosclerosis development.

Lethally irradiated $Ldlr^{-/-}$ mice were injected with bone marrow from $Bcma^{+/+}$ (hem- $Bcma^{+/+}$) or $Bcma^{-/-}$ donors (hem- $Bcma^{-/-}$) and were fed an atherogenic diet for 10 weeks. **a**, $Bcma$ mRNA in the spleen ($n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). **b**, Representative photomicrographs of H&E-stained aortic root lesions (left) and average lesion size in the aortic origin (right) expressed as μm^2 per section ($n = 11$ hem- $Bcma^{+/+}$, $n = 12$ hem- $Bcma^{-/-}$ mice). **c**, Whole body weight, plasma triglyceride levels and en face lesion size ($n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). **d**, Total plasma cholesterol ($n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). **e**, Absolute numbers of FO/T2, MZ, CD21 $^{+}$ CD23 $^{-}$ ($P = 0.028$), T1 ($P = 0.023$), NF ($P = 0.031$) and B-1 B cells ($n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). **f**, Absolute numbers of CD3 $^{+}$, CD4 $^{+}$ and CD8 $^{+}$ T cells ($n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). **g**, **h**, Frequencies of peritoneal B-1a ($P = 0.003$), B-1b ($P = 0.021$) and CD23 $^{+}$ B-2 cells ($P = 0.0003$) (**g**) and total IgM ($P = 0.002$), IgG1, IgG2b ($P = 0.002$), IgG2c ($P = 0.001$), IgG3 ($P = 0.032$) and IgA plasma antibody titers (**h**; $n = 11$ hem- $Bcma^{+/+}$, $n = 13$ hem- $Bcma^{-/-}$ mice). All results show mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (two-tailed Mann–Whitney *U*-test or two-tailed unpaired Student's *t*-test). Scale bar, 200 μm .

Source data

Extended Data Fig. 3 APRIL is produced by mouse and human VSMCs.

a, *TNFSF13* gene expression in human tissues in the Genotype-Tissue Expression (GTEx) project³⁶. The GTEx project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data described in this manuscript were obtained from the GTEx Portal on 21 January 2021 and dbGaP accession number phs000424.v8.p2. The results show median (aorta: median = 55.05, n = 432; coronary artery: median = 40.7, n = 240). **b, c**, Bulk RNA-seq analysis of VSMCs from the aortic arch (AA) and descending thoracic aorta (DT) (**b**; n = 3–5 mice) (GSE117963) and from mouse primary VSMCs that were stored in Trizol after isolation or had been cultured for 4–5 passages until the analysis (**c**; GSE17858). *TNFSF13*, *MYH11* and *KI67* gene expression are depicted. **d**, *TNFSF13* and *IL6* gene expression by human umbilical artery smooth muscle cells that were stimulated in quadruplicate with recombinant human TNF, native human LDL or human oxLDL (TNF stimulation is representative of three independent experiments; P = 0.003). Results show mean \pm s.e.m. ** P < 0.01 (one-way ANOVA and Tukey's test).

Extended Data Fig. 4 APRIL binds HSPG2.

a, Representative photomicrographs (left) and quantification (right) of HUVECs incubated with either human or mouse Flag-APRIL in the presence or absence of heparin and stained with the anti-Flag M2 antibody conjugated to FITC and analysed by confocal microscopy. **b**, Flow cytometry analysis of HUVECs incubated with Flag-APRIL or flag-tagged bacterial alkaline phosphatase (Flag-BAP) and stained with the anti-Flag M2 antibody. **c**, Identification of protein binding partners of APRIL in HUVEC culture by performing a pull-down assay with agarose beads coupled to the anti-Flag M2 antibody followed by MS analysis. **d**, APRIL binding to coated HSPGs from mouse basement membrane quantified by ELISA determined in triplicate (** P < 0.0001, two-tailed unpaired Student *t* test). Data shown are representative of at least two independent

experiments (**a**, **b**, **d**). **e**, **f**, Photomicrographs of mouse carotid artery sections incubated with mouse multimeric Flag–APRIL and stained with either an anti-Flag antibody conjugated to PE or with an anti-mouse APRIL biotinylated antibody (2C8) and streptavidin conjugated to PE (**e**; scale bar, 75 µm, data derived from two independent experiments) or with mouse multimeric Flag–APRIL in the presence or absence of heparin and stained with an anti-APRIL biotinylated antibody (2C8) or with an anti-HSPG2 or only secondary antibody (only 2nd) (**f**; scale bar, 75 µm, data derived from one experiment). **g**, Quantitative surface plasmon resonance (Biacore) analysis of the affinity of soluble human Fc–APRIL (total), human canonical Fc–APRIL (human Fc–c-APRIL), human non-canonical Fc–APRIL (human Fc–nc-APRIL), mouse canonical Fc–APRIL (mouse Fc–c-APRIL), mouse non-canonical Fc–APRIL (mouse Fc–nc-APRIL) and negative controls EDAR–Fc and human Fas–Fc to biotinylated heparin coupled to streptavidin Sensor Chip A ($n = 3$ independent experiments). All results show mean ± s.e.m. IntDen, integrated density.

[Extended Data Fig. 5 Anti-APRIL antibodies 108, 2C8 and Apry-1-1 are specific for mouse APRIL.](#)

a, Coomassie blue analyses of anti-mAPRIL mAb 108 and 2C8 under reducing conditions. **b**, Isotyping of the Fc portions of anti-mAPRIL 108, Apry-1-1 and 2C8. Purified antibodies coated on an ELISA plate were revealed with peroxidase-conjugated antibodies against different isotypes. **c**, Inhibitory activity of 108 and Apry-1-1 compared to that of TACI-Fc on human and mouse APRIL. Flag–human APRIL and two splice variants of Flag–mouse APRIL (\pm Ala112) were titrated on BCMA–Fas reporter cells in the presence of a fixed, non-saturating concentration of 108, Apry-1-1 or TACI-Fc. The data show that 108 and Apry-1-1 inhibit both splice variants of mAPRIL at roughly stoichiometric ratios, but do not cross-react with human APRIL. **d**, ELISA for 2C8 binding to mouse APRIL. Binding of 2C8 to plates coated with human Fc–mouse APRIL was evaluated with a peroxidase-coupled anti-mouse antibody. **a**, **b**, Data are representative of two independent experiments.

[Source data](#)

Extended Data Fig. 6 APRIL competes for binding of LDL to proteoglycans.

a, Representative photomicrographs (left) and quantification (right) of anti-ApoB antibody binding to mouse carotid artery sections incubated with human native LDL in the presence or absence of mouse multimeric Flag-APRIL analysed by both confocal and epifluorescence microscopy (without APRIL, $n = 7$; with APRIL, $n = 9$; $P = 0.0004$). **b**, The amount of bound human LDL (triplicate; quantified by flow cytometry) on the surface of HEK293 wild-type cells in the presence of different amounts of human recombinant APRIL. **c**, Representative photomicrographs of ApoB-stained lesions in the aortic origin (left) and ApoB-positive area normalized to DAPI⁺ lesion area (right) of $Ldlr^{-/-}Tnfsf13^{+/+}$ or $Ldlr^{-/-}Tnfsf13^{-/-}$ mice that were fed an atherogenic diet for 10 weeks ($n = 8$ $Ldlr^{-/-}Tnfsf13^{+/+}$ mice, $n = 12$ $Ldlr^{-/-}Tnfsf13^{-/-}$ mice; $P = 0.035$). Data shown are pooled from four independent experiments with seven to nine sections per group (**a**), representative of three independent experiments (**b**). All results show mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$ (two-tailed unpaired Student's *t*-test). Scale bars, 20 μm (**a**) and 200 μm (**c**).

[Source data](#)

Extended Data Fig. 7 Treatment with the blocking anti-APRIL antibody Apry1-1 does not alter B cells, IgM and plasma lipid levels in *Apoe*^{-/-} mice.

Apoe^{-/-} mice were treated biweekly for 10 weeks with a mixture of anti-mouse APRIL antibody (Apry1-1) and control-Ig (anti-APRIL group), or TACI-Ig and isotype IgG2b (TACI-Ig group), or isotype IgG2b and control-Ig (control group) and were fed an atherogenic diet for the last 8 weeks of the study. **a–f**, Dot plots show the numbers of total splenic B cells (**a**), follicular (FO) B cells (**b**), marginal zone (MZ) B cells (**c**), and frequencies of peritoneal B-1a (**d**), peritoneal B-1b (**e**) and total IgM antibody levels (**f**) in plasma. **g**, Wild-type mice were injected intraperitoneally with either 1 μg mouse multimeric Flag-APRIL or a mixture of 1 μg Flag-APRIL and 10 μg anti-mouse APRIL antibody (Apry-

1-1). The amount of Flag-APRIL in plasma was measured by ELISA one, three and six hours after the injection ($n = 4$ mice Flag-APRIL, $n = 5$ mice Flag-APRIL + anti-APRIL (Apry1-1)). **h**, Whole body weight, plasma triglyceride and cholesterol levels. **a–f, h**, All results show mean \pm s.e.m.; ($n = 10$ *Apoe*^{-/-} control, $n = 12$ *Apoe*^{-/-} TACI-Ig, $n = 10$ *Apoe*^{-/-} anti-APRIL). *** $P < 0.001$, **** $P < 0.0001$ (one-way ANOVA and Newman-Keuls test).

[Source data](#)

[Extended Data Fig. 8 Epitope mapping of anti-human APRIL antibodies, and native canonical and non-canonical APRIL differ in size but are produced by the same gene.](#)

a–g, Epitope mapping of anti-human APRIL antibodies. **a**, The epitopes recognized by Aprily1, 2, 3, 5 and 10 were mapped by western blot of truncated APRIL proteins. **b–d**, Aprily5 and Aprily1 or Aprily2 recognize distinct epitopes (**d**), whereas Aprily3 and Aprily10 recognize epitopes distinct from those of Aprily1, Aprily2 and Aprily5 (**e, f**). **g**, Expression of all constructs was validated by western blot with anti-Fc antibody. **h, i**, Human serum was depleted of APRIL using the anti-human APRIL antibodies Aprily1, Aprily2, Aprily3, Aprily5, Aprily6, Aprily8, Aprily9, Aprily10, Mahya-1, 110.6, the biological atacicept (TACI-Ig: a recombinant fusion protein of the receptor TACI and the Fc region of Ig, that binds to APRIL) or the negative control EctoD1, and then analysed with a c-APRIL-specific (**h**; ELISA 1) or an nc-APRIL-specific ELISA (**i**; ELISA 2). Data are derived from one experiment in this format. **j–n**, Native canonical and non-canonical APRIL differ in size. **j**, Flag-human APRIL (from c-APRIL ELISA 1 standards) was depleted on TACI-Fc (or TNFR2-Fc as control) and/or on Aprily2 (or mIgG1 as control). APRIL was then detected by c-APRIL-specific (top) or nc-APRIL-specific (bottom) ELISA. **k, l**, Flag-human APRIL (from APRIL ELISA 1 standards) was depleted on immobilized TACI-Fc or on Aprily2, and the flow-through was then size-fractionated by size exclusion chromatography (SEC) and detected in fractions by c-APRIL-specific (**k**; ELISA 1) or nc-APRIL-specific (**l**; ELISA 2) ELISA. TACI-Fc and Aprily2 beads used for depletion were then acid-eluted. **m, n**, The neutralized eluate was size-fractionated, and APRIL

in fractions was detected with c-APRIL-specific (**m**) or nc-APRIL-specific (**n**) ELISA. These results indicate that Flag–c-APRIL has the size of a 3-mer, whereas nc-APRIL is much larger. **o**, **p**, Canonical and non-canonical APRIL are produced by the same *TNFSF13* gene locus. The *TNFSF13* gene (which encodes APRIL) was inactivated in human macrophage cell line U937 by CRISPR–Cas9 technology. As a control, the *TNFSF13B* gene (which encodes BAFF) was also deleted. APRIL in supernatants was measured with a c-APRIL-specific (**o**) and an nc-APRIL-specific (**p**) ELISA. 105, 110, 301 and 302 depict different clones.

[Source data](#)

[Extended Data Fig. 9 LC–MS-based parallel reaction monitoring \(PRM\) analysis of tryptic digest of purified human canonical or non-canonical Fc–APRIL.](#)

a–c, Raw data were analysed using Skyline software and extracted product ion chromatograms (XICs) are shown either in the form of peaks (top) or total sum of integrated product ion areas (bottom) for the three selected peptides EEQYNSTYR (Fc part) (**a**), LNLSPHGTFLGFVK (tryptic C terminus APRIL) (**b**) and LNLSPHGTFLGFVKL (mislabeled tryptic C terminus APRIL) (**c**). MS₂ fragment ion spectra for the selected peptide precursor ions are illustrated at bottom right. Although the peptide shown in **a** is representative for comparable injection amounts of canonical versus non-canonical Fc–APRIL, the C-terminal mislabeled full tryptic peptide shown in **c** is undetectable in non-canonical APRIL. Relative abundances are given in arbitrary units. Right, FASTA sequence of Fc–APRIL with selected tryptic peptide sequences highlighted in blue or red. Note the different scales in **b** (10^9) and **c** (10^6). **d**, Structure of human c-APRIL highlighting the importance of the C terminus for the folding of the different forms (canonical and non-canonical) of APRIL. The representation based on protein data bank accession number 1XU1 highlights the last two C-terminal amino acids (Lys232, Leu233). The N-terminal amino acid of the TNF homology domain (His98) and Asp142 are also shown. All of these residues are conserved in mouse APRIL and human APRIL, although the sequence surrounding Asp142 is different in mouse and human. The C-terminal carboxylic group of Leu233 is very

close to His98 of the same monomer (3.6 Å, 4.1 Å and 3.4 Å in the three monomers) and also very close to His98 of the neighbouring monomer (4.3 Å, 3.8 Å and 3.8 Å). Thus, His98 and the carboxylic group of Leu233 seem to form a ring of six salt bridges at the top surface of APRIL. In addition, Lys232 contacts Asp142 (4.3 Å, 4.3 Å and 5.7 Å in the three mouse APRIL monomers), and is only 3.2 Å from Asp142 in human APRIL (PDB accession number 4ZCH).

Extended Data Table 1 Human clinical studies

Supplementary information

Supplementary Figure 1

Quantified purified nc-APRIL and c-APRIL used as standards in nc-APRIL specific and c-APRIL specific ELISA, respectively.

Reporting Summary

Supplementary Table 1

Contains data referring to Extended Data Figure 9a-c.

Supplementary Table 2

Contains data referring to Extended Data Table 1a, b.

Supplementary Table 3

Contains data referring to Extended Data Table 1c.

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. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

[**Source Data Extended Data Fig. 7**](#)

[**Source Data Extended Data Fig. 8**](#)

Rights and permissions

[**Reprints and Permissions**](#)

About this article

Cite this article

Tsiantoulas, D., Eslami, M., Obermayer, G. *et al.* APRIL limits atherosclerosis by binding to heparan sulfate proteoglycans. *Nature* **597**, 92–96 (2021). <https://doi.org/10.1038/s41586-021-03818-3>

- Received: 14 June 2019
- Accepted: 12 July 2021
- Published: 25 August 2021

- Issue Date: 02 September 2021
 - DOI: <https://doi.org/10.1038/s41586-021-03818-3>
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03818-3>

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

- Article
- [Published: 14 July 2021](#)

SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape

- [Tyler N. Starr](#) [ORCID: orcid.org/0000-0001-6713-6904](#)¹,
[Nadine Czudnochowski](#)²,
[Zhuoming Liu](#) [ORCID: orcid.org/0000-0001-8198-0976](#)³,
[Fabrizia Zatta](#)⁴,
[Young-Jun Park](#) [ORCID: orcid.org/0000-0003-2901-6949](#)⁵,
[Amin Addetia](#)¹,
[Dora Pinto](#)⁴,
[Martina Beltramello](#)⁴,
[Patrick Hernandez](#) [ORCID: orcid.org/0000-0001-9704-5093](#)²,
[Allison J. Greaney](#)^{1,6},
[Roberta Marzi](#) [ORCID: orcid.org/0000-0001-6025-6735](#)⁴,
[William G. Glass](#)⁷,
[Ivy Zhang](#)^{7,8},
[Adam S. Dingens](#) [ORCID: orcid.org/0000-0001-9603-9409](#)¹,
[John E. Bowen](#) [ORCID: orcid.org/0000-0003-3590-9727](#)⁵,
[M. Alejandra Tortorici](#) [ORCID: orcid.org/0000-0002-2260-2577](#)⁵,
[Alexandra C. Walls](#) [ORCID: orcid.org/0000-0002-9636-8330](#)⁵,
[Jason A. Wojcechowskyj](#)²,
[Anna De Marco](#)⁴,
[Laura E. Rosen](#) [ORCID: orcid.org/0000-0002-8030-0219](#)²,
[Jiayi Zhou](#) [ORCID: orcid.org/0000-0002-4231-3422](#)²,
[Martin Montiel-Ruiz](#) [ORCID: orcid.org/0000-0001-6200-9578](#)²,
[Hannah Kaiser](#) [ORCID: orcid.org/0000-0002-3991-7401](#)²,
[Josh R. Dillen](#)²,
[Heather Tucker](#)²,
[Jessica Bassi](#)⁴,
[Chiara Silacci-Fregni](#)⁴,

- [Michael P. Housley](#)²,
- [Julia di Julio](#)²,
- [Gloria Lombardo](#)⁴,
- [Maria Agostini](#)²,
- [Nicole Sprugasci](#)⁴,
- [Katja Culap](#) [ORCID: orcid.org/0000-0002-0956-0018](#)⁴,
- [Stefano Jaconi](#) [ORCID: orcid.org/0000-0001-7527-4434](#)⁴,
- [Marcel Meury](#)²,
- [Exequiel Dellota Jr](#)²,
- [Rana Abdelnabi](#)⁹,
- [Shi-Yan Caroline Foo](#) [ORCID: orcid.org/0000-0002-6380-4917](#)⁹,
- [Elisabetta Cameroni](#)⁴,
- [Spencer Stumpf](#)³,
- [Tristan I. Croll](#)¹⁰,
- [Jay C. Nix](#) [ORCID: orcid.org/0000-0002-4041-4975](#)¹¹,
- [Colin Havenar-Daughton](#) [ORCID: orcid.org/0000-0002-2880-3927](#)²,
- [Luca Piccoli](#) [ORCID: orcid.org/0000-0002-1085-6502](#)⁴,
- [Fabio Benigni](#)⁴,
- [Johan Neyts](#) [ORCID: orcid.org/0000-0002-0033-7514](#)⁹,
- [Amalio Telenti](#) [ORCID: orcid.org/0000-0001-6290-7677](#)²,
- [Florian A. Lempp](#) [ORCID: orcid.org/0000-0001-6103-8078](#)²,
- [Matteo S. Pizzuto](#) [ORCID: orcid.org/0000-0001-5776-654X](#)⁴,
- [John D. Chodera](#) [ORCID: orcid.org/0000-0003-0542-119X](#)⁷,
- [Christy M. Hebner](#)²,
- [Herbert W. Virgin](#) [ORCID: orcid.org/0000-0001-8580-7628](#)^{2,12,13},
- [Sean P. J. Whelan](#)³,
- [David Veesler](#) [ORCID: orcid.org/0000-0002-6019-8675](#)⁵,
- [Davide Corti](#) [ORCID: orcid.org/0000-0002-5797-1364](#)⁴,
- [Jesse D. Bloom](#) [ORCID: orcid.org/0000-0003-1267-3408](#)^{1,6,14} &
- [Gyorgy Snell](#) [ORCID: orcid.org/0000-0003-1475-659X](#)²

Nature volume 597, pages 97–102 (2021)

- 57k Accesses
- 1 Citations
- 542 Altmetric
- [Metrics details](#)

Subjects

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

Abstract

An ideal therapeutic anti-SARS-CoV-2 antibody would resist viral escape^{1,2,3}, have activity against diverse sarbecoviruses^{4,5,6,7}, and be highly protective through viral neutralization^{8,9,10,11} and effector functions^{12,13}. Understanding how these properties relate to each other and vary across epitopes would aid the development of therapeutic antibodies and guide vaccine design. Here we comprehensively characterize escape, breadth and potency across a panel of SARS-CoV-2 antibodies targeting the receptor-binding domain (RBD). Despite a trade-off between in vitro neutralization potency and breadth of sarbecovirus binding, we identify neutralizing antibodies with exceptional sarbecovirus breadth and a corresponding resistance to SARS-CoV-2 escape. One of these antibodies, S2H97, binds with high affinity across all sarbecovirus clades to a cryptic epitope and prophylactically protects hamsters from viral challenge. Antibodies that target the angiotensin-converting enzyme 2 (ACE2) receptor-binding motif (RBM) typically have poor breadth and are readily escaped by mutations despite high neutralization potency. Nevertheless, we also characterize a potent RBM antibody (S2E12⁸) with breadth across sarbecoviruses related to SARS-CoV-2 and a high barrier to viral escape. These data highlight principles underlying variation in escape, breadth and potency among antibodies that target the RBD, and identify epitopes and features to prioritize for therapeutic development against the current and potential future pandemics.

[Download PDF](#)

Main

The most potently neutralizing antibodies to SARS-CoV-2—including those in clinical use¹⁴ and dominant in polyclonal sera^{15,16}—target the spike RBD. Mutations in the RBD that reduce binding by antibodies have emerged among SARS-CoV-2 variants^{17,18,19,20,21}, highlighting the need for antibodies and vaccines that are robust to viral escape. We have previously described S309⁴, an antibody that exhibits potent effector functions and neutralizes all current SARS-CoV-2 variants^{22,23} and the divergent sarbecovirus SARS-CoV-1. S309 forms the basis for an antibody therapy (VIR-7831, recently renamed sotrovimab) that has received emergency use authorization from the US Food and Drug Administration for treatment of COVID-

19²⁴. Longer term, antibodies with broad activity across sarbecoviruses would be useful to combat potential future spillovers⁶. These efforts would be aided by a systematic understanding of the relationships between antibody epitope, resistance to viral escape, and breadth of sarbecovirus cross-reactivity. Here we address this question by comprehensively characterizing a diverse panel of antibodies, including S309, using deep mutational scanning, pan-sarbecovirus binding assays, in vitro selection of viral escape, and biochemical and structural analyses.

Antibody potency, escapability and breadth

We identified a panel of anti-SARS-CoV-2 antibodies with distinct properties (Fig. [1a](#), Extended Data Table [1](#)), including six antibodies newly described in this study. These antibodies bind different epitopes within the RBM and the non-RBM ‘core’ of the RBD. The antibody panel spans a range of neutralization potencies and binding affinities (Fig. [1a](#), Extended Data Fig. [1a–c](#)).

Fig. 1: Potency, escapability, and breadth of a panel of RBD antibodies.

 **figure1**

a, SARS-CoV-2 neutralization potency (authentic virus ($n = 3$) and spike-pseudotyped VSV particles (VSV pp) ($n = 3$ to 8) on Vero E6 cells), Fab–RBD binding affinities measured by surface plasmon resonance (SPR) ($n = 2$ to 4), and epitope classifications. Additional details in Extended Data Table 1. Asterisks indicate antibodies that are newly described in this paper. Epitope classes are described in refs. 9,15. IC₅₀, half-maximal inhibitory concentration; K_D , dissociation constant; NA, not applicable. **b, c**, Left, deep mutational scanning maps of mutations that escape binding

by antibodies targeting the core RBD (**b**) or the RBM (**c**). Letter height indicates the strength of the mutation's escape from antibody binding. Letters coloured by effect on folded RBD expression (**b**) or ACE2 binding affinity²⁶ (**c**). Right, relative functional epitope size and escapability, scaling from 0 (no escape mutations) to 1 (largest epitope and most escapable antibody). Bottom, heat maps illustrate variability among sarbecoviruses or SARS-CoV-2 sequences. **d**, Antibody binding to a pan-sarbecovirus RBD panel. Heat map illustrates binding from fluorescence-activated cell sorting (FACS)-based selections (colour bar, bottom left). Asterisks indicate reduced-affinity binding in secondary binding assays (Extended Data Fig. [4a–f](#)).

We used deep mutational scanning to map how all amino-acid mutations in the SARS-CoV-2 RBD affect binding by each antibody³ (Fig. [1b,c](#), Extended Data Fig. [2](#)). Some antibodies have narrowly focused functional epitopes (the set of residues in which mutations abolish binding²⁵), with binding-escape mutations at just a few key residues (for example, S309, S2D106), whereas others have wider functional epitopes (for example, S2H13; Fig. [1b,c](#)). We previously measured how all RBD mutations affect folded RBD expression and ACE2 binding affinity²⁶ (Fig. [1b,c](#), [letter colours](#)). We used the combined measures of how mutations affect antibody binding and RBD function to compute the ‘escapability’ of each antibody, which reflects the extent to which mutations that escape antibody binding are functionally tolerated (Fig. [1b,c](#), Extended Data Fig. [3a,b](#)). We also investigated the sensitivity of each antibody to mutations among SARS-CoV-2 sequences reported in GISAID (Fig. [1b,c](#), heat map, Extended Data Fig. [3c](#)), and found that some antibodies are more affected by natural SARS-CoV-2 mutations than others, including mutations found in SARS-CoV-2 variants of concern^{27,28,29} (Extended Data Fig. [1d](#)).

We next extended our deep mutational scanning platform to measure binding of each antibody to a pan-sarbecovirus panel of 45 RBDs (Fig. [1d](#), Extended Data Fig. [4a–f](#)). The four antibodies that bind the core RBD exhibit cross-reactive binding to RBDs from SARS-CoV-1 and related ACE2-dependent bat sarbecoviruses, and from sarbecoviruses isolated in Europe and Africa. S304 and S2H97 antibodies also bind RBDs of the most divergent clade from Asia, which has an average of 64% amino acid identity with SARS-CoV-2. S2H97 exhibits notably tight binding to all RBDs tested (Fig. [1d](#), Extended Data Fig. [4f](#)), making it, to our knowledge, the broadest pan-sarbecovirus RBD antibody described to date. Antibodies that bind epitopes within the RBM exhibit more limited cross-reactivity, typically binding only SARS-CoV-2 and the closely related GD-pangolin-CoV RBD. S2E12 stands out among the RBM antibodies we evaluated, as it also binds the RaTG13 and GX-pangolin-CoV RBDs, showing that even within the evolutionarily plastic RBM^{19,26} there are epitopes that enable greater breadth than others.

The pan-sarbecovirus S2H97 antibody

To understand the structural basis for cross-reactive sarbecovirus binding, we determined the structures of the S2H97 antigen-binding fragment (Fab) (X-ray crystallography, 2.65 Å resolution), S2X35 Fab (X-ray crystallography, 1.83 Å resolution) and S2E12 Fab (X-ray crystallography, 2.95 Å resolution) bound to SARS-CoV-2 RBD (Fig. 2a, Extended Data Table 2). This panel of cross-reactive antibodies emphasizes the core RBD as a general target of broad antibody binding owing to its conservation among sarbecoviruses, reflected in the diverse core RBD surfaces targeted by the broadest of these antibodies (Fig. 2a, Extended Data Fig. 5a–g).

Fig. 2: The pan-sarbecovirus S2H97 antibody.

 figure2



a, Composite model of the SARS-CoV-2 trimer with cross-reactive antibodies. Epitopes recognized by each Fab are shown as coloured surface and ACE2 footprint is shown as a black outline. **b**, Integrative features of the S2H97 structural footprint (5 Å cut-off). Heat maps illustrate evolutionary variability (blue), functional constraint from previous deep mutational scans (grey) and energetic contribution to binding from the static crystal structure (predicted energy) or molecular dynamics (MD) simulation (green). Logo plot as in Fig. 1b. Asterisk, introduction of N-linked glycosylation motifs. **c**, S2H97 breadth of neutralization (spike-pseudotyped VSV on Vero E6 cells).

Curves representative of at least two independent experiments. Data are mean \pm s.d. from three technical replicates. Geometric mean IC₅₀ is shown. **d**, S2H97 neutralization of SARS-CoV-2 variants (Extended Data Fig. 1d; spike-pseudotyped VSV on Vero E6 cells). Points represent individual measurements and bars show mean fold-change in neutralization potency relative to the Wuhan-1 strain with the D614G S mutation. **e**, Negative-stain electron microscopy imaging of native-like soluble prefusion S trimer alone (left) or incubated with S2H97 (right). Micrographs are representative of 51 (SARS-CoV-2 S) and 173 (S + S2H97) micrographs. **f**, S2H97 prophylactic efficacy in Syrian hamsters. Median infectious virus titres (left, TCID50) and RNA levels (right) 4 days after SARS-CoV-2 challenge in hamsters prophylactically dosed with 25 mg kg⁻¹ S2H97 or isotype control. The double daggers indicate two hamsters with undetectable S2H97 levels (below 50 ng ml⁻¹) at the time of viral challenge. ** $P = 0.0048$ (virus titre) and $P = 0.0048$ (RNA) vs control, two-sided Mann–Whitney test (hamsters with no detectable serum antibody excluded). **g**, Blockade of binding¹⁵ by sera from SARS-CoV-2-infected (top) or vaccinated (bottom) individuals. BD₈₀ is the concentration that blocks 80% of binding. Percentages of samples with blockade above the lower detection limit are indicated. WT, wild type.

[Source data](#)

The exceptionally cross-reactive S2H97 antibody targets a previously undescribed cryptic antigenic site, which we designated site V (Fig. 2a, b). S2H97 binding is facilitated by packing of the heavy chain complementarity-determining region 3 (CDR3) into an RBD crevice at the centre of the epitope, together with polar contacts with all three heavy chain CDRs and the light chain CDR2 (Extended Data Fig. 5f). Molecular dynamics simulation of the S2H97 Fab–RBD complex highlights the durability of many of these interactions (Fig. 2b). The surface bound by S2H97 is constrained by the deleterious effects of mutations on folded RBD expression²⁶ (Fig. 2b), and this constraint is probably enhanced by quaternary packing with the N-terminal domain in the closed spike trimer (Extended Data Fig. 6a). Consistent with the conservation of the S2H97 epitope, S2H97 neutralizes diverse sarbecoviruses (Fig. 2c, Extended Data Fig. 4g) and SARS-CoV-2 variants (Fig. 2d).

To understand the evolution of S2H97 breadth, we measured breadth of binding by its germline form, S2H97_{GL}, in which we reverted the 13 somatic mutations (Extended Data Fig. 4h, i). S2H97_{GL} bound all tested sarbecovirus RBDs and exhibited particularly high affinity for SARS-CoV-2-related RBDs. Somatic mutations enhanced affinity across all sarbecoviruses by two orders of magnitude. This general increase in affinity together with the conservative amino acid replacements among paratope residues suggests that framework mutations may contribute to a general improvement in S2H97 binding affinity.

To characterize the mechanism of S2H97 neutralization, we determined a cryo-electron microscopy structure of S2H97 bound to SARS-CoV-2 S (Extended Data Fig. 5*i–l*, Extended Data Table 3). S2H97 binding requires extensive opening of the RBD to unmask its cognate epitope (Extended Data Fig. 6*b*), even more than is required to access the cryptic antigenic site II¹⁵. Similar to other antibodies that only bind the open RBD^{30,31}, S2H97 induces rapid and premature refolding of spike into the post-fusion state (Fig. 2*e*), promotes S1 shedding of cell-surface-expressed spike (Extended Data Fig. 6*c*), and induces a low level of syncytia formation among spike-expressing cells (Extended Data Fig. 6*d*). S2H97 does not interfere with ACE2 binding (Extended Data Fig. 6*e*). Like other non-ACE2-competitive antibodies^{31,32}, S2H97 neutralization is attenuated in cells that overexpress ACE2 (Extended Data Fig. 6*f*). Consistent with its ability to neutralize spike-mediated viral entry, S2H97 inhibits spike-mediated cell-cell fusion (Extended Data Fig. 6*g*). Together, these experiments suggest that the mechanism of neutralization by S2H97 involves receptor-independent conversion of S to the post-fusion state³⁰, thereby inhibiting ACE2-mediated cell entry.

Next, we determined the prophylactic efficacy of S2H97 *in vivo* using a Syrian hamster model of infection. We administered hamsters with 25 mg kg⁻¹ S2H97 2 days before intranasal challenge with SARS-CoV-2 and assessed viral RNA load and infectious viral titres in the lungs 4 days after infection. S2H97 prophylaxis reduced the number of RNA copies by more than 10,000-fold relative to control in the four hamsters that had detectable circulating antibody levels at the time of challenge, and reduced infectious viral titres to the lower detection limit in these hamsters (Fig. 2*f*). The two hamsters not exhibiting a reduction in viral load had circulating S2H97 levels below the limit of quantification (50 ng ml⁻¹) at the time of viral challenge (Extended Data Fig. 6*h*), which may reflect a failure in the intraperitoneal administration procedure. Therefore, S2H97 demonstrates that antibodies to the newly identified antigenic site V can be protective *in vivo*.

Finally, we performed serum blockade of binding experiments¹⁵, demonstrating that antibodies competing with S2H97 binding are rare in infection- and vaccine-elicited sera (Fig. 2*g*). This sub-dominance of antigenic site V may be explained by the inaccessibility of the epitope as illustrated in the cryoEM structure. However, the protective nature and exceptional breadth of S2H97 suggests that updated immunogen designs, such as those based on the RBD^{33,34,35}, could unmask antigenic site V to better elicit S2H97-like antibodies.

Breadth and escapability among RBM epitopes

Our survey reveals variation in the escapability and breadth of antibodies that target the RBM (Fig. 1*c,d*), which is immunodominant (Fig. 2*g*) but variable over

sarbecovirus and SARS-CoV-2 evolution. We performed in vitro selection experiments to identify spike-expressing vesicular stomatitis virus (VSV) mutants that emerge in the presence of each of seven monoclonal antibodies (Fig. 3a, Extended Data Fig. 7a, b) to further understand escape from these antibodies.

Fig. 3: Breadth and escapability among RBM antibodies.



a, Escape mutations in spike-expressing VSV passaged in the presence of monoclonal antibody. Plot shows effects of indicated mutations on antibody (x-axis) and ACE2 (y-axis) binding²⁶. Mutations are classed according to whether they are accessible by single mutation from the Wuhan-1 strain (single-nt accessible) and those that require multi-nucleotide mutations. **b**, Neutralization of SARS-CoV-2 variants by S2E12 (spike-pseudotyped VSV on Vero E6 cells), relative to Wuhan-1 D614G, as in Fig. 2d.

c, S2E12 breadth of neutralization (spike-pseudotyped VSV on 293T-ACE2 cells). Points represent mean of biological duplicates. **d**, Replicative fitness of S2E12 escape mutations identified in **a** on Vero E6 cells. Data are mean \pm s.e.m. from triplicate experiments. **e, f**, Structures of S2E12 Fab (**e**) and S2D106 Fab (**f**) bound to SARS-CoV-2 RBD. RBD sites coloured by escape (colour bar, centre). The E484 side chain is included for visualization purposes only but was not included in the final S2D106-bound structure owing to weak density. Max, maximum. **g, h**, Integrative features of the structural footprints (5 Å cut-off) of S2E12 (**g**) and S2D106 (**h**). Sites are ordered by the frequency of observed mutants among SARS-CoV-2 sequences on GISAID. Heat maps as in Fig. [2b](#). Logo plots as in Fig. [1c](#), but showing only amino acid mutations accessible via single-nucleotide mutation from Wuhan-1 strain for comparison with **a**. Bar plots illustrate frequency of SARS-CoV-2 mutants and their validated effects on antibody neutralization (spike-pseudotyped VSV on Vero E6 cells). Red indicates greater than tenfold increase in IC₅₀ due to mutation. ND, not determined.

Many RBM antibodies such as S2X58 and S2D106 select mutations present in SARS-CoV-2 variants of concern (for example, L452R and E484K)[27,28,29](#). By contrast, S2E12 selects viral mutants at sites that do not exhibit substantial variation among circulating SARS-CoV-2, and S2E12 correspondingly neutralizes a diverse panel of SARS-CoV-2 variants^{[17](#)} (Fig. [3b](#)). S2E12 is also unique in its breadth among RBM antibodies (Fig. [1d](#)), neutralizing VSV pseudotyped with each of the four SARS-CoV-2 clade sarbecovirus spikes (Fig. [3c](#), Extended Data Fig. [4j](#)). As with S2H97, somatic mutations in S2E12 enhanced affinity across sarbecoviruses, though the increase in affinity was more modest than for S2H97 (Extended Data Fig. [4k,l](#)).

Conservation of the S2E12 epitope among SARS-CoV-2 variants could reflect the relative rarity of S2E12-like antibodies in polyclonal sera leading to little antigenic pressure at these sites (Fig. [2g](#)), together with functional constraint in the S2E12 epitope (escapability being the lowest for S2E12 and S2H97 among the 12 antibodies evaluated). Indeed, the strong antibody-escape mutations that emerged in S2E12 viral escape selections decrease ACE2 binding affinity^{[26](#)} (Fig. [3a](#)) and reduce replicative fitness in a bulk competition experiment between spike-expressing VSV variants passaged in the absence of antibody (Fig. [3d](#)).

To understand the structural basis for the unique breadth and robustness of S2E12 to escape, we compared its structure to that of S2D106 Fab (cryoEM, 4.0 Å resolution local refinement) bound to SARS-CoV-2 RBD (Fig. [3e,f](#), Extended Data Fig. [5g,h](#), [m-p](#), Extended Data Tables [2,3](#)). We also integrated evolutionary, functional, and structural details for the sites in each antibody's structural footprint (Fig. [3g,h](#)). S2E12 and S2D106 bind the receptor-binding ridge, with 8 residues shared between their footprints. S2E12 binding is oriented towards extensive packing of the ACE2-contact

residue F486_{RBD} within a cavity lined by aromatic residues at the antibody light–heavy-chain interface (Fig. 3e, Extended Data Fig. 5g), as was seen with the homologous antibody COV2-2196³⁶. Sites within the S2E12 footprint that exhibit less functional constraint (for example, E484 and S477) are located at the periphery of the interface, explaining the robustness of S2E12 towards SARS-CoV-2 variants (Fig. 3b, g). This structural interface also explains the breadth of S2E12 towards RaTG13 and GX-pangolin-CoV (Fig. 1d), as the F486L mutation present in these sarbecoviruses retains the central hydrophobic packing.

In contrast to S2E12, S2D106 binding is centred on residue E484_{RBD} which may form a salt bridge with R96_{LC}, in addition to nonpolar contacts between F490_{RBD} and residues in the heavy chain CDR2 (Fig. 3f, Extended Data Fig. 5h). Although the long heavy chain CDR3 packs intimately across the surface of the RBD, there are no crucial CDRH3–RBD contacts that are sensitive to mutation. S2D106 escape is therefore highly focused on E484 and F490, which are functionally tolerant and exhibit variation among SARS-CoV-2 sequences (Fig. 3h). This comparison between S2E12 and S2D106 highlights how small differences in the RBD–antibody interface affect the breadth and robustness of each antibody to viral escape.

The landscape of RBD epitopes

Last, we examined how escapability, breadth, and neutralization potency relate to one another and to the RBD epitope. We used our binding-escape maps (Fig. 1b, c), together with comparable maps published for other RBD antibodies^{3,20,21,36,37}, to project antibodies into a two-dimensional space on the basis of similarities in sites of binding-escape mutations (Fig. 4a).

Fig. 4: Antibody epitope, potency, breadth and escapability.

 **figure4**

a, Multidimensional scaling projection of similarities in antibody binding-escape maps from the present (red) and previous (grey) studies. Pie charts illustrate the RBD subdomains where mutations confer escape (key, bottom left: RBM, residues 437–508; ACE2 contact, 4 Å cutoff). Structural projections of escape arrayed around the

perimeter (colour bar, bottom right), with grey outlines tracing structural footprints. **b–d**, Projected epitope space from **a** annotated by antibody properties: SARS-CoV-2-neutralizing potency (**b**), sarbecovirus-binding breadth (**c**) and escape by natural SARS-CoV-2 mutants (**d**). For each property, antibodies are coloured such that purple reflects the most desirable antibody (colour bar, right): most potent neutralization (\log_{10} scale), highest breadth, and lowest natural frequency of escape mutants (\log_{10} scale). **e**, Relationship between SARS-CoV-2 neutralization potency and sarbecovirus breadth for antibodies described in this work and in a parallel study³⁷ (S2X259). **f**, Relationship between functional epitope size and SARS-CoV-2 RBD binding affinity. **g**, Relationship between natural SARS-CoV-2 escape mutant frequency (Extended Data Fig. [3c](#)) and sarbecovirus breadth.

We annotated our projection of epitope space using antibody properties such as in vitro neutralization potency, breadth and escapability (Fig. [4b–d](#), Extended Data Fig. [7c,d](#)). The most potently neutralizing antibodies (for example, S2E12 and S2D106) bind epitopes in the RBM, whereas antibodies targeting the core RBD are less potently neutralizing (Fig. [4b](#)). Of note, RBD antibodies can protect *in vivo* through other mechanisms beyond neutralization^{12,13,22}. Antibodies with broad sarbecovirus binding target the core RBD (Fig. [4c](#)). Our panel therefore extends previous observations^{4,5,32,38} to highlight a general trade-off between sarbecovirus breadth and potency of SARS-CoV-2 neutralization (Fig. [4e](#)). Nonetheless, some cross-reactive antibodies exhibit intermediate in vitro neutralization potency (for example, S309 and S2X259³⁷), and the highly potent RBM-directed antibody S2E12 exhibits modest breadth, highlighting the existence of antibodies that balance neutralization potency and breadth.

The size of an antibody's functional epitope (Fig. [1b,c](#)) is not strongly influenced by the structural location of the epitope (Extended Data Fig. [7c](#))—instead, narrower functional epitopes are associated with higher Fab–RBD binding affinity (Fig. [4f](#)). However, an antibody's escapability, which integrates how escape mutations affect RBD folding and ACE2 affinity, is influenced by variation in these functional constraints across the RBD structure. For example, antibodies that cluster with S2E12 exhibit lower escapability (Extended Data Fig. [7c](#)) and frequency of natural SARS-CoV-2 escape mutants (Fig. [4d](#)). As highlighted by S2E12 and S2H97 above, antibodies with even a modest degree of sarbecovirus breadth show a greatly reduced frequency of escape mutations among circulating SARS-CoV-2 variants (Fig. [4g](#)).

Principles for antibody and vaccine development

Ongoing SARS-CoV-2 evolution^{19,27,28,29}, long-term antigenic evolution of other human coronaviruses^{39,40} and the spillover potential of diverse sarbecovirus

lineages^{6,7} indicate the importance of developing antibodies and vaccines that are robust to viral evolution. In this work, we identify antibody and epitope features that can guide this process. Although *in vitro* neutralization potency is often prioritized for lead selection, our results suggest that this will bias antibodies towards RBM epitopes, many of which are poorly conserved in the short-term evolution of SARS-CoV-2¹⁹ and the long-term evolution of sarbecoviruses⁷. Our results suggest that additional prioritization of high-affinity binding and at least a moderate degree of sarbecovirus breadth will yield antibodies with improved resistance to viral escape^{4,5}.

A long-term goal is to develop antibodies and vaccines that cross-react with distant sarbecovirus lineages capable of zoonotic spillover. We have identified a cryptic epitope capable of eliciting pan-sarbecovirus immunity, represented by S2H97. Though S2H97-like antibodies are rare in polyclonal sera, the protective capacity and exceptional breadth of S2H97 indicates that pan-sarbecovirus vaccines could seek to improve responses to this epitope by unmasking this and other cryptic broadly neutralizing epitopes^{5,37,41}. Broader cross-reactivity among betacoronavirus lineages including MERS and OC43 has been reported for antibodies that bind the spike S2 domain^{32,38,42}. Though S2H97 breadth does not extend beyond sarbecoviruses, its discovery expands our view of what can be achieved via a potent RBD-directed antibody response.

The global emergence of variants of concern (VOC) has been an important feature of the pandemic^{27,28,29}. Mutations in VOC occur in immunodominant RBM epitopes (for example, residues E484, K417 and L452) and influence binding by polyclonal serum and some therapeutic antibodies^{17,18,19,20,21}. We cannot predict which mutations will next rise to prominence as SARS-CoV-2 continues to evolve, but it seems likely that they will include additional RBM mutations that affect recognition by infection- and vaccine-elicited antibodies^{1,2,15,16,19}. Therefore, antibody discovery efforts focused on breadth^{4,5}, aided by high-resolution differentiation among antibody epitopes as generated here, can inform the development of antibody and vaccine countermeasures with greater robustness to immune escape in the current SARS-CoV-2 pandemic and utility for potential future sarbecovirus spillovers.

Methods

Mammalian cell lines

Cell lines were received from ATCC (Vero E6, Vero, BHK-21, CHO-K1, HEK293T/17), Takara (Lenti-X 293T) and Thermo Fisher Scientific (ExpiCHO-S, Expi293F and Freestyle 293-F). MA104 cells were a gift from H. B. Greenberg (Stanford School of Medicine). The 293T-ACE2 cells are described in references^{31,43}.

Vero and MA104 cell lines tested negative for mycoplasma contamination. Other cell lines were not tested. No authentication was performed beyond manufacturer standards.

Isolation of peripheral blood mononuclear cells, plasma and sera

Samples from three individuals who had recovered from SARS-CoV-2, designated as donors S2H (age 36, male), S2D (age 70, male) and S2X (age 52, male) were obtained under study protocols approved by the local Institutional Review Board (Canton Ticino Ethics Committee, Switzerland). All donors provided written informed consent for the use of blood and blood components (such as peripheral blood mononuclear cells (PBMCs), sera or plasma). Blood drawn from donor S2X was obtained at day 48 (S2X16, S2X35 and S2X58 antibodies) and 75 (S2X227) after symptoms onset. Blood from donor S2H was obtained at day 17 (S2H13 and S2H14), day 45 (S2H58) and day 81 (S2H97) after symptoms onset. Blood from donor S2D was obtained at day 98 (S2D106) after symptoms onset.

PBMCs were isolated from blood draw performed using tubes pre-filled with heparin, followed by Ficoll density gradient centrifugation. PBMCs were either used fresh for SARS-CoV-2 Spike protein-specific memory B cell sorting or stored in liquid nitrogen for later use. Sera were obtained from blood collected using tubes containing clot activator, followed by centrifugation and storage at -80 °C.

Sera for blockade of binding serological assays were obtained from 3 cohorts of SARS-CoV-2 convalescent (average age 52, range 25–78, 55% male) or vaccinated (average age 49, range 28–69, 65% male) individuals under study protocols approved by the local Institutional Review Boards (Canton Ticino Ethics Committee, Switzerland, the Ethical Committee of Luigi Sacco Hospital, Milan, Italy, and WCG North America, Princeton, NJ, USA). All donors provided written informed consent for the use of blood and blood components (such as PBMCs, sera or plasma) and were recruited at hospitals or as outpatients.

B-cell isolation and recombinant monoclonal antibody production

The discovery and initial characterization of six of the antibodies in our panel has been previously reported (S309 and S304, refs. [4,15](#); S2X35, S2H13 and S2H14, ref. [15](#); and S2E12, ref. [8](#)), and the other six (S2H97, S2X16, S2H58, S2D106, S2X58 and S2X227) are described for the first time in this article. Starting from freshly isolated PBMCs or upon cell thawing, B cells were enriched by staining with CD19 PE-Cy7 (BD Bioscience 557835, 1:50) and incubation with anti-PE MicroBeads (Miltenyi Biotec 130-048-801, 1:100), followed by positive selection using LS columns (Miltenyi Biotec). Enriched B cells were stained with anti-IgM (BioLegend 314508,

1:20), anti-IgD (BD Bioscience 555779, 1:40), anti-CD14 (BD Bioscience 562691, 1:50) and anti-IgA (Southern Biotech 2050-09, 1:400), all PE labelled, and prefusion SARS-CoV-2 S with a biotinylated Avi-tag (in house produced) conjugated to Streptavidin–Alexa-Fluor 647 (Life Technologies S21374, 1:40). SARS-CoV-2 S-specific IgG⁺ memory B cells were sorted by flow cytometry via gating for PE negative and Alexa-Fluor 647 positive cells. Cells were cultured for the screening of positive supernatants. Antibody VH and VL sequences were obtained by PCR with reverse transcription (RT–PCR) and monoclonal antibodies were expressed as recombinant human Fab fragment or as IgG1 (G1m3 allotype) carrying the half-life-extending M428L/N434S (LS) mutation in the Fc region. ExpiCHO-S cells (Thermo Fisher Scientific) were transiently transfected with heavy and light chain expression vectors as previously described⁴. Affinity purification was performed on Äkta Xpress FPLC (Cytiva) operated by Unicorn software version 5.11 (Build 407) using HiTrap Protein A columns (Cytiva) for full length human monoclonal antibodies and CaptureSelect CH1-XL MiniChrom columns (Thermo Fisher Scientific) for Fab fragments, using PBS as mobile phase. Buffer exchange to the appropriate formulation buffer was performed with a HiTrap Fast desalting column (Cytiva). The final products were sterilized by filtration through 0.22-μm filters and stored at 4 °C.

Using the Database IMGT (<http://www.imgt.org>), the Vh and Vl germline gene family and the number of somatic mutations were determined by analysing the homology of the Vh and Vl sequences to known human *V*, *D* and *J* genes. Germline-reverted sequences of the Vh and Vl were constructed using IMGT/V-QUEST. The S2E12 and S2H97 germline-reverted antibodies (G1m17 allotype) were produced by ATUM. Germline-reverted Fab fragments were generated by digestion of the corresponding IgGs.

Epitope classes shown in Figs. 1a, 2g are defined as in Piccoli et al. ¹⁵. In brief, the classification of these epitope classes results from Octet binning experiments using structurally characterized antibodies, structural insights to define the recognition of open-only RBD and ability of antibodies to interfere with RBD binding to ACE2. In particular, site Ia is accessible only in the open state of RBD and largely overlaps the ACE2 footprint; site Ib is accessible in both open and closed RBD states and overlaps in part the ACE2 footprint; site IIa is in the core RBD (accessible only in the open RBD state) and antibodies binding to this site interfere with binding to ACE2, site IIc is also in the core RBD but targeted by antibodies that do not interfere with binding to ACE2; site IV is fully accessible on both open and closed RBDs and is defined by the footprint of the S309 antibody.

Neutralization of authentic SARS-CoV-2 by entry-inhibition assay

Neutralization was determined using SARS-CoV-2-Nluc, an infectious clone of SARS-CoV-2 (based on strain 2019-nCoV/USA_WA1/2020) which encodes nanoluciferase in place of the viral ORF7 and demonstrates comparable growth kinetics to wild-type virus⁴⁴. Vero E6 cells (ATCC, CRL-1586) were seeded into black-walled, clear-bottom 96-well plates at 2×10^4 cells per well and cultured overnight at 37 °C. The next day, 9-point fourfold serial dilutions of monoclonal antibodies were prepared in infection media (DMEM + 10% FBS). SARS-CoV-2-Nluc was diluted in infection media at a final multiplicity of infection (MOI) of 0.01 plaque-forming units (PFU) per cell, added to the monoclonal antibody dilutions and incubated for 30 min at 37 °C. Medium was removed from the Vero E6 cells and monoclonal antibody–virus complexes were added and incubated at 37 °C for 24 h. Media was removed from the cells, Nano-Glo luciferase substrate (Promega) was added according to the manufacturer’s recommendations, incubated for 10 min at room temperature and the luciferase signal was quantified on a Victor Nivo plate reader (Perkin Elmer).

SARS-CoV-2 spike pseudotyped VSV generation and neutralization assay

Replication-defective VSV pseudoviruses⁴⁵ expressing SARS-CoV-2 spike protein were generated as previously described⁴⁶ with some modifications. Plasmids encoding SARS-CoV-2 spike single-mutant variants were generated by site-directed mutagenesis of the wild-type plasmid, pcDNA3.1(+)-spike-D19⁴⁷, and plasmids encoding multiply mutated SARS-CoV-2 variants of concern were generated using a multistep overlap extension PCR protocol^{23,48}, in which sequential, overlapping fragments were designed to introduce all mutations, which were PCR assembled and cloned into the pcDNA3.1 vector using the Takara In-fusion HD cloning kit following manufacturer’s instructions.

Lenti-X 293T (Takara, 632180) cells were seeded in 10-cm dishes at a density of 1×10^5 cells per cm² and the following day transfected with 5 µg of spike expression plasmid with TransIT-Lenti (Mirus, 6600) according to the manufacturer’s instructions. For the neutralization assays with variants of concern (Figs. 2d, 3b), Lenti-X 293T cells were seeded in 10-cm dishes at a density of 5×10^6 cells per cm², and transfected the following day with 10 µg of spike expression plasmid. One day after transfection, cells were infected with VSV (G*ΔG-luciferase) (Kerafast, EH1020-p.m.) for 1 h, rinsed three times with PBS, then incubated for an additional 24 h in complete medium at 37 °C. The cell supernatant was clarified by centrifugation, filtered (0.45 µm), aliquoted, and frozen at –80 °C.

For VSV pseudovirus neutralization assays, Vero E6 cells (ATCC, CRL-1586) were grown in DMEM supplemented with 10% FBS and seeded into clear bottom white 96

well plates (Costar, 3903) at a density of 2×10^4 cells per well. The next day, monoclonal antibodies were serially diluted in pre-warmed complete media, mixed at a 1:1 ratio with pseudovirus and incubated for 1 h at 37 °C in round bottom polypropylene plates. Media from cells was aspirated and 50 µl of virus–monoclonal antibody complexes were added to cells and then incubated for 1 h at 37 °C. An additional 100 µl of prewarmed complete media was then added on top of complexes and cells incubated for an additional 16–24 h. Conditions were tested in duplicate wells on each plate and at least six wells per plate contained uninfected, untreated cells (mock) and infected, untreated cells ('no mAb control').

Medium containing virus–monoclonal antibody complexes was then aspirated from cells and 100 µl of a 1:4 dilution of Bio-glo (Promega, G7940) in PBS was added to cells. For neutralization assays with variants of concern, 50 µl of a 1:2 dilution of SteadyLite Plus (Perkin Elmer) in PBS with Ca²⁺ and Mg²⁺ was added to cells in place of Bio-glo. Plates were incubated for 10 min at room temperature and then were analysed on an Envision plate reader (PerkinElmer), or for variants of concern assays, a Synergy H1 Hybrid Multi-Mode reader (Biotek).

Relative light unit (RLU) measurements for infected wells were subtracted by the average of RLU values for the mock wells (background subtraction) and then normalized to the average of background subtracted 'no mAb control' RLU values within each plate. Per cent neutralization was calculated by subtracting from 1 the normalized monoclonal antibody infection condition. Data were analysed and visualized with Prism (version 8.4.3). IC₅₀ values were calculated from the interpolated value from the log(inhibitor) versus response–variable slope (four parameters) nonlinear regression with an upper constraint of <100. Neutralization experiments with wild-type SARS-CoV-2 S and single-mutant variants were conducted on three independent days, that is, biological replicates, where each biological replicate contains a technical duplicate. IC₅₀ values across biological replicates are presented as geometric mean. The loss or gain of neutralization potency across spike variants was calculated by dividing the variant IC₅₀ by the parental IC₅₀ within each biological replicate. Neutralization experiments with SARS-CoV-2 S variants of concern were conducted in biological duplicates, with IC₅₀ values normalized by the corresponding wild-type measurement, and presented as arithmetic mean of the duplicate experiments.

SARS-CoV-2 spike pseudotyped VSV neutralization on 293T-ACE2 cells

To investigate the effect of ACE2 expression on S2H97 neutralization, Vero E6 cells were seeded at 20,000 cells per well in black clear-bottom 96-well plates. 293T-ACE2 cells³¹ were seeded at 35,000 cells per well in black clear-bottom 96-well plates that

had been pre-coated with poly-d-lysine (Gibco). The next day, SARS-CoV-2 spike-pseudotyped VSV neutralizations with S2E12, S309 and S2H97 were performed as described above. Neutralizations were performed in triplicate wells.

Sarbecovirus spike pseudotyped VSV neutralization by S2H97

Mammalian expression constructs (pcDNA3.1(+) or pTwist-CMV) encoding the spike proteins from various sarbecoviruses with a C-terminal deletion of 19 amino acids (D19) were synthesized for SARS-CoV-2 (GenBank: QOU99296.1), SARS-CoV-1 Urbani (GenBank: AAP13441.1), hCoV-19/pangolin/Guangdong/1/2019 (GD-pangolin-CoV, GenBank: QLR06867.1), pangolin coronavirus Guanxi-2017 (GX-pangolin-CoV, GenBank: QIA48623.1), and bat sarbecovirus WIV1 (WIV1, GenBank: AGZ48828.1). Lenti-X 293T cells (Takara, 632180) were seeded in 15-cm dishes such that the cells would reach 80% confluence after culturing overnight. The following day, cells were transfected using TransIT-Lenti (Mirus, 6600) according to the manufacturer's instructions. One day after transfection, cells were infected with VSV (G* Δ G-luciferase) (Kerafast, EH1020-p.m.). The supernatant containing sarbecovirus pseudotyped VSV was collected 2 days after transfection, centrifuged at 1,000g for 5 min, aliquoted and frozen at -80 °C.

For neutralization assays, cells supporting robust pseudovirus infection were seeded into clear bottom white-walled 96-well plates at 20,000 cells per well in 100 µl culture media. Vero E6 cells were used for VSV-SARS-CoV-2, VSV-SARS-CoV-1, and VSV-GD-pangolin-CoV. BHK-21 cells (ATCC, CCL-10) stably expressing ACE2 were used for VSV-GX-pangolin-CoV and VSV-WIV1. After culturing cells overnight, 1:3 serial dilutions of antibody were prepared in DMEM in triplicate. Pseudovirus was diluted in DMEM and added to each antibody dilution such that the final dilution of pseudovirus was 1:20. Pseudovirus–antibody mixtures were incubated for 1 h at 37 °C. Medium was removed from the cells and 50 µl of pseudovirus:antibody mixtures were added. One hour after infection, 50 µl of culture medium was added to wells containing pseudovirus:antibody mixtures and incubated overnight at 37 °C. Medium was then removed and 100 µl of 1:1 diluted DPBS:Bio-Glo (Promega, G7940) luciferase substrate was added to each well. The plate was shaken at 300 rpm at room temperature for 10 min after which RLU reading were taken on an EnSight (Perkin Elmer) microplate reader. Per cent neutralization was determined by first subtracting the mean background (cells with luciferase substrate alone) RLU values of 6 wells per plate for all data points. Per cent neutralization for each antibody concentration was calculated relative to no-antibody control wells for each plate. Per cent neutralization data were analysed and graphed using Prism (GraphPad, v9.0.1). Absolute IC₅₀ values were calculated by fitting a curve using a nonlinear regression model (variable slope, 4 parameters) and values were interpolated from the curve at y = 50%. The geometric

mean from at least two independent experiments was calculated using Excel (Microsoft, version 16.45).

Sarbecovirus spike pseudotyped VSV neutralization by S2E12

Spikes from SARS-CoV-2 (CAD0240757.1), RaTG13 (QHR63300.2), GD-pangolin (QLR06867.1), GX-pangolin (QIA48623.1), SARS-CoV-1 Tor2 (YP009825051), WIV1 (AGZ48831.1) and WIV16 (ALK02457.1) were used to pseudotype VSV. To produce pseudotyped viruses, HEK293T/17 (ATCC, CRL-11268) seeded in 10-cm dishes in DMEM supplemented with 10% FBS, 1% penicillin-streptomycin were transfected with plasmids using lipofectamine 2000 (Life Technologies) following manufacturer's instructions. One day post-transfection, cells were infected with VSV ($G^*\Delta G$ -luciferase) for 2 h and washed 4 times with DMEM, before adding medium supplemented with anti-VSV-G antibody (I1-mouse hybridoma supernatant at 1:50 dilution, from CRL-2700, ATCC). Pseudotyped particles were collected 18 h after inoculation, clarified by centrifugation at 2,000g for 5 min, concentrated 10 \times with a 30 kDa cut-off membrane filter, and stored at -80 °C. For S2E12 neutralization experiments, 293T cells stably expressing ACE2 (BEI #NR-52511)⁴³ in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin were seeded at 40,000 cells per well in clear-bottom white-walled 96-well plates and cultured overnight at 37 °C. Twelve threefold serial dilutions of S2E12 antibody were prepared in DMEM, and antibody dilutions were mixed 1:1 with pseudotyped VSV in the presence of 1:50 diluted anti-VSV-G antibody. After 45 min incubation at 37 °C, 40 μ l of antibody-virus mixture was added to cells, and 40 μ l DMEM was added 2 h post-infection. After 17–20 h, 50 μ l One-Glo-EX substrate (Promega) was added to the cells. Cells were incubated in the dark for 5–10 min before luminescence reading on a Varioskan LUX plate reader (Thermo Fisher Scientific). Relative luciferase unit values were converted to percentage of neutralization and plotted with a nonlinear regression curve fit in GraphPad Prism. Measurements were performed in duplicate with two independent productions of pseudotyped virus.

Recombinant protein production

SARS-CoV-2 RBD wild-type protein for SPR binding assays (with N-terminal signal peptide and C-terminal thrombin cleavage site-TwinStrep-8 \times His-tag) was expressed in Expi293F (Thermo Fisher Scientific) cells at

37 °C and 8% CO₂. Transfections were performed using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Cell culture supernatants were collected three days after transfection and supplemented with 10× PBS to a final concentration of 2.5× PBS (342.5 mM NaCl, 6.75 mM KCl and 29.75 mM phosphates). SARS-CoV-2 RBDs were purified using 1 or 5 ml HisTALON Superflow cartridges (Takara Bio) and subsequently buffer exchanged into 1× HBS-N buffer (Cytiva) or PBS using a Zeba Spin Desalting (Thermo Fisher Scientific) or HiPrep 26/10 (Cytiva) desalting column.

SARS-CoV-2 RBD wild-type for crystallization (with N-terminal signal peptide, ETGT and C-terminal 8×His-tag) was expressed similarly as described above in the presence of 10 µM kifunensine. Cell culture supernatant was collected four days after transfection and supplemented with 10× PBS to a final concentration of 2.5× PBS. Protein was purified using a 5 ml HisTALON Superflow cartridge followed by size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl. For crystallization of the RBD-S2X259-S2H97 and RBD-S2E12-S304-S309 Fab complexes, RBD was deglycosylated by overnight incubation with EndoH glycosidase at 4 °C.

RBDs from other sarbecoviruses for SPR (with N-terminal signal peptide and C-terminal thrombin cleavage site-TwinStrep-8×His-tag) were expressed in Expi293F cells at 37 °C and 8% CO₂. Cells were transfected using PEI MAX (Polysciences) at a DNA:PEI ratio of 1:3.75. Transfected cells were supplemented three days after transfection with 3 g l⁻¹ glucose (Bioconcept) and 5 g l⁻¹ soy hydrolysate (Sigma-Aldrich). Cell culture supernatant (423 ml) was collected 7 days after transfection and supplemented with 47 ml 10× binding buffer (1 M Tris-HCl, 1.5 M NaCl, 20 mM EDTA, pH 8.0) and 25 ml BioLock (IBA) and incubated on ice for 30 min. Proteins were purified using a 5-ml Strep-Tactin XT Superflow high-capacity cartridge (IBA) followed by buffer exchange to PBS using HiPrep 26/10 desalting columns (Cytiva).

Prefusion-stabilized SARS-CoV-2 spike proteins for SPR (residues 14-1211, either D614 or D614G), containing the 2P and Furin cleavage site

mutations⁴⁹ with a mu-phosphatase signal peptide and a C-terminal Avi-8×His-C-tag or C-terminal 8xHis-Avi-C-tag were expressed in Freestyle 293-F cells (Thermo Fisher Scientific, R79007) at 37 °C and 8% CO₂. Transfections were performed using 293fectin as a transfection reagent. Cell culture supernatant was collected after three days and purified over a 5 ml C-tag affinity matrix. Elution fractions were concentrated and injected on a Superose 6 Increase 10/300 GL column (Cytiva) with 50 mM Tris-HCl pH 8.0 and 200 mM NaCl as running buffer.

SARS-CoV-2 HexaPro spike protein for cryoEM analysis was produced in Freestyle 293-F cells grown in suspension using FreeStyle 293 expression medium (Life Technologies) at 37 °C in a humidified 8% CO₂ incubator rotating at 130 rpm. The cultures were transfected using PEI (9 µg ml⁻¹) with cells grown to a density of 2.5 million cells per ml and cultivated for three days. The supernatants were collected and cells resuspended for another three days, yielding two collections from each such culture. Spike proteins were purified from clarified supernatants using a 5 ml cobalt affinity column (Cytiva, HiTrap TALON crude), concentrated and flash frozen in a buffer containing 20 mM Tris pH 8.0 and 150 mM NaCl before analysis.

SARS-CoV-2 S native-like ectodomain trimer for refolding assays was engineered with a mu-phosphatase signal peptide beginning at 14Q, a mutated S1/S2 cleavage site (SGAR), and a TEV cleavage, fold-on trimerization motif, and 8x His tag appended to the C terminus (K1211). Native-like spike was expressed and purified as described for SARS-CoV-2 HexaPro spike above.

Recombinant hACE2 for SPR (residues 19–615 from Uniprot Q9BYF1 with a C-terminal AviTag-10×His-GGG-tag, and N-terminal signal peptide) was produced by ATUM. Protein was purified via Ni Sepharose resin followed by isolation of the monomeric hACE2 by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) pre-equilibrated with PBS.

SPR binding assays

SPR binding measurements were performed using a Biacore T200 instrument with CM5 sensor chip covalently immobilized with StrepTactin XT to capture recombinant RBD proteins (data in Fig. 1a, Extended Data Figs. 1c, 4f, i, l). Running buffer was Cytiva HBS-EP+ (pH 7.4). All measurements were performed at 25 °C. Fab (or hACE2) analyte concentrations were 11, 33, 100 and 300 nM, run as single-cycle kinetics. Double reference-subtracted data were fit to a 1:1 binding model using Biacore T200 Evaluation (version 3.1) or Biacore Insight Evaluation (version 2.0.15) software. K_D values above 1 μM were determined from fits where the maximum SPR signal at saturation (R_{\max}) was set as a constant determined from results for higher affinity analytes binding to the same RBD at the same surface density. Data where averages are not given are representative of duplicate or triplicate measurements (except measurements with germline Fabs, which were singleton measurements).

To corroborate the SARS-CoV-2 RBD binding measurements, experiments were also performed in two additional formats, both with monovalent analytes (data in Extended Data Table 1): (1) Fab binding to SARS-CoV-2 spike ectodomain was measured using CM5 sensor chips immobilized with anti-AviTag pAb (Genscript, A00674-40) for capturing S, other experiment parameters same as above, and (2) RBD binding to IgG was measured using CM5 sensor chips immobilized with anti-human Fc pAb (Southern Biotech, 2014-01) for capturing IgG, with RBD analyte concentrations of 3.1, 12.5, and 50 nM, other experiment parameters same as above. Fit results yield an apparent K_D for the spike-binding experiments because the kinetics also reflect spike conformational dynamics. Spike ectodomain was D614G with C-terminal 8xHis-Avi-C-tag for all measurements except S2X58 binding was performed with D614 spike with C-terminal Avi-8 \times His-C-tag. For the comparison of mature and germline-reverted antibody binding to RaTG13, the data reported are from experiment format (2) with IgG as ligand. These data (and others indicated in Extended Data Table 1 as 'biphasic kinetics') were fit to a heterogeneous ligand model, due to an artefactual kinetic phase with very slow dissociation that often arises when RBD is an analyte; the lower affinity of the two K_D values reported by the fit is given as the K_D (the two K_D values are separated by at least one order of magnitude).

Deep mutational scanning mutant escape profiling

We used a previously described deep mutational scanning approach³ to comprehensively identify RBD mutations that escape binding by each antibody. This approach leverages duplicate RBD mutant libraries²⁶, which contain virtually all of the 3,819 possible amino acid mutations in the background of the Wuhan-Hu-1 RBD sequence. Library variants were previously linked to short identifier barcode sequences and sorted to purge the library of variants that strongly decrease ACE2 binding affinity or expression of folded RBD³.

We first used an isogenic yeast strain expressing the unmutated SARS-CoV-2 RBD and flow cytometry to identify the EC90 of each antibody's binding to yeast-displayed SARS-CoV-2 RBD. We then performed library selections as previously described^{3,20}, labelling libraries with the EC90 concentration of antibody to standardize escape mutation sensitivity across selections. In brief, libraries of yeast were induced for surface expression, washed, and labelled with the primary antibody for 1 h at room temperature. Cells were washed, and secondarily labelled with 1:200 PE-conjugated goat anti-human-IgG antibody (Jackson ImmunoResearch 109-115-098) to label for bound antibody, and 1:100 FITC-conjugated chicken anti-Myc-tag (Immunology Consultants Lab, CYMC-45F) to label for RBD surface expression. We prepared controls for setting FACS selection gates by labelling yeast expressing the unmutated SARS-CoV-2 RBD with the same antibody concentration as library selections (1×), 100× reduced antibody concentration to illustrate the effect of mutations with 100×-reduced affinity, and 0 ng ml⁻¹ antibody to illustrate complete loss of antibody binding. Representative selection gates are shown in Extended Data Fig. 2b. Gates were set and sorting performed with FACSDiva software (version 6.1.3). We sorted approximately 7.5×10^6 RBD⁺ cells per library on a BD FACSAria II, collecting yeast cells from the antibody-escape sort bin (fractions of library falling into antibody escape bin given in Extended Data Fig. 2c). Sorted cells were recovered overnight, plasmids were extracted from the pre-sort and antibody-escape populations, and variant-identifier barcode sequences were PCR amplified and sequenced on an Illumina HiSeq 2500^{3,26}.

As previously described^{3,20}, sequencing counts pre- and post-selection were used to estimate the ‘escape fraction’ for each library variant, which reflects the fraction of yeast expressing a variant that fall into the antibody-escape FACS bin. In brief, we used the dms_variants package (https://jbloomlab.github.io/dms_variants/, version 0.8.2) to process Illumina sequences into variant counts pre- and post-selection using the barcode/RBD variant lookup table from Starr et al. ²⁶. We then computed per-variant escape fraction of variant v (E_v) as: $\frac{(\{E\}_{v} - F) \times N_{post}}{N_{pre}}$, where F is the total fraction of the library that escapes antibody binding (Extended Data Fig. 2c), $\{E\}_{v}$ are the sequencing counts of variant v in the RBD library after and before FACS selection, respectively (with a pseudocount of 0.5 added to all counts), and N_{post} and N_{pre} are the total counts of all variants after and before FACS selection, respectively. We then applied computational filters to remove variants with low pre-sort sequencing counts or highly deleterious mutations that might cause artefactual antibody escape due to global unfolding or loss of expression of RBD on the cell surface. Specifically, we filtered out variants whose pre-selection sequencing counts were lower than the 99th percentile counts of variants containing premature stop codons, which were largely purged by the prior sorts for RBD expressing and ACE2-binding RBD variants. We also removed variants with ACE2 binding scores below -2.35 or RBD expression scores below -1 , and variants containing individual mutations with effects below these thresholds, using the variant- and mutation-level deep mutational scanning measurements of Starr et al. ²⁶. We also filtered out rare mutations with low coverage in the libraries, retaining mutations that were sampled on at least one single-mutant barcoded variant or at least two multiply mutated variants in each replicate. Last, to decompose single-mutation escape fractions for each antibody, we implemented global epistasis models⁵⁰ using the dms_variants package to estimate the effect of each individual amino acid mutation, exactly as described in ref. ²⁰.

Antibody escape selections were conducted in full duplicate using independently generated and assayed SARS-CoV-2 mutant libraries (see correlations in Extended Data Fig. 2e,f). The reported escape fractions throughout the paper are the average across the two replicates, and these final per-mutation escape fractions are provided on GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs/blob/main/results/supp_data/vir_antibodies_raw_data.csv). Interactive visualizations of antibody escape maps (https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Vir_mAbs) were created using dms-view⁵¹.

Sarbecovirus library binding assays

A curated set of all unique sarbecovirus RBD amino acid sequences was gathered, including the sarbecovirus RBD sequence set reported by Letko et al.⁷, along with additional unique RBD sequences among SARS-CoV-1 epidemic strains reported by Song et al.⁵², BtKY72⁵³ and new sarbecovirus sequences RmYN02⁵⁴, GD-pangolin-CoV (consensus RBD reported in figure 3a of Lam et al.⁵⁵), and GX-pangolin-CoV⁵⁵ (P2V, ambiguous nucleotide within codon 515 (SARS-CoV-2 spike numbering) resolved to retain F515, which is conserved in all other sarbecoviruses). A list of all RBDs and sequence accession numbers is available on GitHub (https://github.com/jbloomlab/SARSr-CoV_RBD_MAP/blob/main/data/RBD_accessions.csv).

To define clades of sarbecovirus RBDs, an alignment of amino acid RBD sequences was generated using mafft⁵⁶ with gap opening penalty 4.5 (alignment available on GitHub: https://github.com/jbloomlab/SARSr-CoV_RBD_MAP/blob/main/data/RBD_aa_aligned.fasta). The corresponding nucleotide sequence alignment was generated from the amino acid alignment using PAL2NAL⁵⁷. The gene sequence phylogeny was inferred using RAxML version 8.2.12⁵⁸, with the GTRGAMMA substitution model and a partition model with separate parameters for first, second and third codon positions. The Hibecovirus RBD sequence Hp-Zhejiang2013 (GenBank: KF636752) was used as an outgroup for rooting of the sarbecovirus phylogeny.

All unique sarbecovirus RBD protein-coding sequences were ordered from IDT, Twist or Genscript, and cloned into our yeast display vector²⁶. Sequences were pooled and appended with downstream 16-nt barcode sequences according to the protocol described in Starr et al. ²⁶. Long-read circular consensus sequences spanning the 16-nt barcode and RBD genotype were gathered on a PacBio Sequel v2.0 and processed exactly as described in Starr et al. ²⁶. This yielded a barcode:variant lookup table for the sarbecovirus RBD library analogous to that used for SARS-CoV-2 mutant libraries. This table is available on GitHub (https://github.com/jbloomlab/SARSr-CoV_RBD_MAP/blob/main/data/barcode_variant_table.csv).

The pooled sarbecovirus RBD library was labelled, sorted, and quantified as described for the SARS-CoV-2 mutant libraries above, except we only sorted around 1 million RBD⁺ cells owing to the reduced library size. Sequencing and quantification of per-variant antibody escape was conducted as described above. Data for the HKU3-8 RBD are not shown, as this RBD was not successfully expressed in our yeast-display platform. For several antibodies, we performed a secondary experiment, selecting the sarbecovirus RBD library with a more stringent ‘full escape’ gate to select out only variants exhibiting complete loss of binding (Extended Data Fig. [2b,c](#)).

For follow-up quantitative binding assays, select sarbecovirus RBDs were cloned into the yeast-display platform as isogenic stocks. Binding assays were conducted across a titration series of antibody in 96-well plates, and binding at each antibody concentration (geometric mean fluorescence intensity in the PE channel among RBD⁺ (FITC⁺) cells) was determined via flow cytometry and fit to a four-parameter Hill curve to identify the EC₅₀ (midpoint).

Analysis of mutations in natural SARS-CoV-2 sequences

All spike sequences on GISAID⁵⁹ as of 2 May 2021, were downloaded and aligned using mafft⁵⁶. Sequences from non-human origins, sequences with gaps or ambiguous characters in the RBD, and sequences with more than

eight amino acid differences from the Wuhan-Hu-1 reference sequence (GenBank MN908947, residues N331-T531) were removed. We determined mutation frequencies compared with Wuhan-Hu-1 reference from this final alignment of 1,190,241 sequences. We acknowledge all contributors to the GISAID EpiCoV database for their sharing of sequence data. The contributors to GISAID EpiCoV are listed at
https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs/blob/main/data/gisaid_hcov-19_acknowledgement_table_2021_03_04.pdf.

Quantitative summary metrics of antibody properties

The relative epitope size of an antibody was calculated as the sum of per-mutant escape fractions that are at least five times the global median escape fraction (to minimize the impact of variation in background noise on the summation). For this summation, escape fractions were normalized to the maximum per-mutation escape fraction, to account for slight variation in the largest per-mutation escape fraction measured between selections.

The relative escapability of an antibody was calculated the same as relative epitope size, but each mutation was multiplied by two weighting factors scaled from 0 to 1 that reflect the impact of that mutation on ACE2-binding affinity and RBD expression as measured in our prior deep mutational scan²⁶. The relationship between weighting factors and mutation effect on each property is shown in Extended Data Fig. 3a. Mutations with <-1 effect on either property are effectively zeroed out in the escapability summation. Mutations with effects between -1 and 0 have intermediate weights, and mutations with 0 or positive effects are assigned weight factors of 1.

Antibody susceptibility to escape by natural SARS-CoV-2 mutations was calculated as the summed GISAID frequencies of all escape mutations, where escape mutations (all labels in Extended Data Fig. 3c) are defined as those with escape fraction greater than five times the median escape fraction as above. These summed natural escape frequencies are tabulated in the plot headers in Extended Data Fig. 3c.

The summary breadth of an antibody was calculated from the sarbecovirus RBD library escape selection using the standard gating (Extended Data Fig. 4b), only. Although we have various follow-up binding data illustrating reduced affinity binding for some ‘escaped’ sarbecovirus RBDs, these follow-up experiments were not conducted systematically for all antibody–RBD combinations, and therefore would bias breadth estimates. Breadth of binding was calculated as the frequency of all sarbecovirus RBDs that are bound with affinity within the FACS selection gating threshold, weighted by clade representation. Breadth was normalized to give equal representation to each of the four sarbecovirus clades to account for different depth of sampling. Within the SARS-CoV-1 clade, all human 02/03 strains and civet and human 03/04 strains were similarly down-weighted to each represent 1/8 of the possible breadth within the SARS-CoV-1 clade (together with the six bat sarbecoviruses in this clade). As an example, breadth for S304 is calculated as [4/4 + ([6/6] + [6/6] + 5)/8 + 2/2 + 0/21]/4 = 0.72, based on the data shown in Extended Data Fig. 4b.

Multidimensional scaling projection of antibody epitopes

Multidimensional scaling projection in Fig. 4 was performed using the Python scikit-learn package. We first computed the similarity and dissimilarity in the sites of escape between each pair of antibodies, exactly as described in Greaney et al.³, and performed metric multidimensional scaling with two components on the matrix of dissimilarities between all antibody pairs. Antibodies in this layout were coloured with pie charts proportional to the total squared site-wise escape that falls into the labelled structural regions (RBM, residues 437 to 508, ACE2 contact defined as 4 Å cut-off based on 6M0J crystal structure⁶⁰, and core RBD otherwise). In this layout, we included all of our previously published antibodies for which we have performed escape mapping via this same approach. These antibodies and their citations include: S2X259³⁷; LY-CoV555²¹; COV2-2196 and COV2-2130³⁶; REGN10933, REGN10987 and LY-CoV016²⁰; and all other COV2 antibodies and CR3022³.

For Fig. 4b–d and Extended Data Fig. 7c, we coloured the antibodies within this layout according to various antibody properties. When appropriate, we

also coloured these previously assayed antibodies, as described below. Extended Data Fig. 7d and the scatter plots in Fig. 4e–g show the relationships between properties for antibodies specifically in this study (and S2X259) for the most direct comparability.

Antibody neutralization potencies illustrated in Fig. 4b incorporate the authentic SARS-CoV-2 neutralization IC₅₀ values as reported in this study (Fig. 1a), together with the live SARS-CoV-2 neutralization IC₅₀ values for the COV2 antibodies reported by Zost et al. 10. We acknowledge that it is imperfect to compare neutralization potencies reported from different labs on different antibody batches, though in this case, both sets are indeed neutralization potencies with authentic virus. We therefore do not directly compare these two sets of measurements in a quantitative manner, but we do note that their joint inclusion in Fig. 4b supports the dichotomy between neutralization potency of core RBD versus RBM antibodies which is supported by either neutralization panel alone.

Sarbecovirus breadth illustrated in Fig. 4c incorporates the pan-sarbecovirus breadth measurements reported in the current study together with more limited breadth measurements for antibodies reported in previous publications. These previously published experiments determined binding within a more restricted sarbecovirus RBD set present in our libraries (SARS-CoV-2, RaTG13, GD-pangolin, SARS-CoV-1 (Urbani), LYRa11 and WIV1). We calculated breadth from this incomplete sarbecovirus sequence set for comparison, but note that these antibodies are limited to a relative breadth of 0.5 because no RBDs from the Africa/Europe or non-ACE2-using Asia clades were included. However, as with neutralization, inclusion of these antibodies nonetheless emphasizes the core RBD–RBM dichotomy in sarbecovirus breadth established by our primary panel.

For illustrations of epitope size and escapability in Fig. 4d and Extended Data Fig. 7c, we calculated these quantities for our previously profiled antibodies as described above. We excluded the antibodies profiled in Greaney et al. 3, as these assays were performed on a prior version of our SARS-CoV-2 mutant library that exhibited different quantitative features of absolute escape, complicating its quantitative comparison to extent of

escape for antibodies profiled in this and our other studies, which all use the same library.

Structural mappings around the perimeter of Fig. 4a were created by mapping total site-wise escape to the B-factor column of PDB structures. Footprints were defined as residues within a 5 Å cut-off of antibody heavy atoms. Structures used were those described in this paper, or previously published structures: ACE2-bound RBD (6M0J)⁶⁰, CR3022-bound RBD (6W41)⁶¹, REGN10987- and REGN10933-bound RBD (6XDG)⁶², CB6-(LY-CoV016) bound RBD (7C01)⁶³, and S304, S309 and S2H14-bound RBD (7JX3)¹⁵.

RBD ELISA

Ninety-six half-area well-plates (Corning, 3690) were coated over-night at 4 °C with 25 µl of sarbecoviruses RBD proteins at 5 µg ml⁻¹ in PBS pH 7.2. Plates were blocked with PBS 1% BSA (Sigma-Aldrich, A3059) and subsequently incubated with serial dilutions of monoclonal antibody for 1 h at room temperature. After 4 washing steps with PBS 0.05% Tween 20 (PBS-T) (Sigma-Aldrich, 93773), goat anti-human IgG secondary antibody (Southern Biotech, 2040-04) was added and incubated for 1 h at room temperature. Plates were then washed 4 times with PBS-T and 4-nitrophenyl phosphate (pNPP, Sigma-Aldrich, 71768) substrate was added. After 30 min of incubation, absorbance at 405 nm was measured by a plate reader (Biotek) and data were plotted using GraphPad Prism.

Binding to cell surface-expressed sarbecovirus S proteins by flow cytometry

ExpiCHO-S cells were seeded at 6×10^6 cells per ml in a volume of 5 ml in a 50-ml bioreactor. Spike coding plasmids were diluted in cold OptiPRO SFM, mixed with ExpiFectamine CHO Reagent (Life Technologies) and added to the cells. Transfected cells were then incubated at 37 °C with 8% CO₂ with an orbital shaking speed of 120 rpm (orbital diameter of 25 mm) for 42 h. Transiently transfected ExpiCHO-S cells were collected and washed two times in wash buffer (PBS, 1% BSA, 2 mM EDTA). Cells were

counted and distributed into round bottom 96-well plates (Corning) and incubated with $10 \mu\text{g ml}^{-1}$ S2H97, S2X35 or S309 monoclonal antibody. Alexa Fluor647-labelled Goat Anti-Human IgG secondary Ab (Jackson ImmunoResearch 109-607-003) was prepared at $1.5 \mu\text{g ml}^{-1}$ and added to cells after two washing steps. Cells were then washed twice and resuspended in wash buffer for data acquisition on a ZE5 cytometer (Bio-Rad).

Crystallization, data collection, structure determination and analysis

To form RBD–Fab complexes for crystallization, SARS-CoV-2 RBD was mixed with a 1.3-fold molar excess of each Fab and incubated on ice for 20–60 min. Complexes were purified on a Superdex 200 Increase 10/300 GL column (Cytiva) preequilibrated with 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. Crystals of the RBD–Fab complexes were obtained at 20 °C by sitting-drop vapour diffusion.

For the SARS-CoV-2 RBD–S2X35–S309 complex, a total of 200 nl complex at 5.4 mg ml^{-1} was mixed with 100 nl mother liquor solution containing 1.85 M ammonium sulfate, 0.1 M Tris pH 8.17, 0.8% (w/v) polyvinyl alcohol, 1% (v/v) 1-propanol, and 0.01 M HEPES pH 7. Crystals were flash frozen in liquid nitrogen using the mother liquor solution supplemented with 20% glycerol for cryoprotection. Data were collected at Beamline 9-2 of the Stanford Synchrotron Radiation Lightsource facility in Stanford, CA and processed with the XDS software package (version 31 January 2020)⁶⁴ to 1.83 Å in space group C222. The RBD–S2X35–S309 Fab complex structure was solved by molecular replacement using phaser⁶⁵ from a starting model consisting of SARS-CoV-2 RBD–S309 Fab (PDB 7JX3) and a homology model for the S2X35 Fab built using the Molecular Operating Environment (MOE) software package from the Chemical Computing Group (<https://www.chemcomp.com>).

For the SARS-CoV-2 RBD-S2H97-S2X259 Fab complex, 200 nl complex at 5.7 mg ml^{-1} was mixed with 200 nl mother liquor solution containing 0.12 M monosaccharides mix, 20% (v/v) ethylene glycol, 10% (w/v) PEG 8000, 0.1 M Tris (base)/bicine pH 8.5, 0.02 M sodium chloride, 0.01 M

MES pH 6 and 3% (v/v) Jeffamine ED-2003. Crystals were flash frozen in liquid nitrogen. Data were collected at Beamline 9-2 of the Stanford Synchrotron Radiation Lightsource facility in Stanford, CA. Data were processed with the XDS software package (version Jan 31, 2020)⁶⁴ for a final dataset of 2.65 Å in space group *P*2₁. The RBD–S2H97–S2X259 Fab complex structure was solved by molecular replacement using phaser from a starting model consisting of SARS-CoV-2 RBD (PDB 7JX3) and homology models for the S2H97 and S2X259 Fabs built using the MOE software package.

For the SARS-CoV-2 RBD–S2E12–S304–S309 Fab complex, 200 nl complex at 4.5 mg ml⁻¹ was mixed with 100 nl of 0.09 M phosphate/citrate pH 5.5, 27% (v/v) PEG smear low, 4% (v/v) polypropylene glycol 400 and 0.02 M imidazole pH 7 or 100 nl of 0.09 M phosphate/citrate pH 5.5, 27% (v/v) PEG smear low, 0.01 M potassium/sodium phosphate pH 7, 1% (v/v) PPGBA 230 and 1.5% (v/v) PPGBA 400. Crystals were flash frozen in liquid nitrogen. Data were collected at the Molecular Biology Consortium beamline 4.2.2 at the Advanced Light Source synchrotron facility in Berkeley, CA. Datasets from two crystals from the two conditions were individually processed and then merged with the XDS software package (version 31 January 2020)⁶⁴ for a final dataset of 2.93 Å in space group *I*4₁22. The RBD–S2E12–S304–S309 Fab complex structure was solved by molecular replacement using phaser from starting models consisting of SARS-CoV-2 RBD–S304–S309 Fab (PDB 7JX3) and S2E12 (PDB 7K3Q).

For all structures, several subsequent rounds of model building and refinement were performed using Coot (version 0.9.5)⁶⁶, ISOLDE (ChimeraX version 1.1/ISOLDE version 1.1)⁶⁷, Refmac5 (version 5.8.0267)⁶⁸ and MOE (version 2019.0102) (<https://www.chemcomp.com>), to arrive at the final models. For all complexes, epitopes on the RBD protein were determined by identifying all RBD residues within a 5.0 Å distance from any Fab atoms. The analysis was performed using the MOE software package and the results were manually confirmed.

Cryo-electron microscopy

SARS-CoV-2 HexaPro S⁶⁹ at 1.2 mg ml⁻¹ was incubated with 1.2-fold molar excess of recombinantly purified S2D106 or S2H97 at 4 °C before application onto a freshly glow discharged 2.0/2.0 UltrAuFoil grid (200 mesh). Plunge freezing used a vitrobot MarkIV (Thermo Fisher Scientific) using a blot force of 0 and 6.5 s blot time at 100% humidity and 23 °C.

For the S/S2D106 dataset, data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. Automated data collection was carried out using Leginon⁷⁰ at a nominal magnification of 130,000× with a pixel size of 0.525 Å. The dose rate was adjusted to 8 counts per pixel per s, and each movie was acquired in super-resolution mode fractionated in 50 frames of 200 ms. A total of 2,166 micrographs was collected, with a defocus range between -0.5 and -2.5 μm. Movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking, and extraction were carried out using Warp⁷¹. Particle images were extracted with a box size of 800 binned to 400 pixels² yielding a pixel size of 1.05 Å.

For the S/S2H97 dataset, data were acquired on an FEI Glacios transmission electron microscope operated at 200 kV equipped with a Gatan K2 Summit direct detector. Automated data collection was carried out using Leginon⁷⁰ at a nominal magnification of 36,000× with a pixel size of 1.16 Å. The dose rate was adjusted to 8 counts per pixel per s, and each movie was acquired in counting mode fractionated in 50 frames of 200 ms. A total of 3,138 micrographs was collected in a single session with a defocus range comprised between -0.5 and -3.0 μm. Preprocessing was performed using Warp⁷¹ and particle images were extracted with a box size of 400 pixels².

For the S/S2D106 and S/S2H97 datasets, two rounds of reference-free 2D classification were performed using CryoSPARC to select well-defined particle images⁷². These selected particles were subjected to two rounds of 3D classification with 50 iterations each (angular sampling 7.5° for 25 iterations and 1.8° with local search for 25 iterations), using our previously reported closed SARS-CoV-2 S structure as initial model⁴⁹ (PDB 6VXX) in

Relion⁷³. 3D refinements were carried out using non-uniform refinement⁷⁴ along with per-particle defocus refinement in CryoSPARC. Selected particle images were subjected to the Bayesian polishing procedure⁷⁵ implemented in Relion3.0 before performing another round of non-uniform refinement in CryoSPARC followed by per-particle defocus refinement and again non-uniform refinement.

To further improve the density of the S2D106 Fab, the particles were then subjected to focus 3D classification without refining angles and shifts using a soft mask on RBD and Fab variable domains with a tau value of 60 in Relion. Particles belonging to classes with the best resolved local density were selected and subject to local refinement using CryoSPARC. Local resolution estimation, filtering, and sharpening were carried out using CryoSPARC. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) of 0.143 criterion and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution⁷⁶. UCSF Chimera⁷⁷ and Coot⁷⁸ were used to fit atomic models into the cryoEM maps. Spike-RBD/S2D106 Fab model was refined and relaxed using Rosetta using sharpened and unsharpened maps⁷⁹.

S2H97-induced spike refolding

Ten micromolar native-like SARS-CoV-2 S was incubated with 13 µM S2H97 Fab for 1 h at room temperature. Samples were diluted to 0.01 mg ml⁻¹ immediately before adsorption to glow-discharged carbon-coated copper grids for ~30 s before a 2% uranyl formate staining. Micrographs were recorded using the Leginon software⁷⁰ on a 120 kV FEI Tecnai G2 Spirit with a Gatan Ultrascan 4000 4k × 4k CCD camera at 67,000× nominal magnification. The defocus ranged from -1.0 to -2.0 µm and the pixel size was 1.6 Å.

Cell-surface antibody-mediated S1 shedding

CHO-K1 cells stably expressing the prototypic SARS-CoV-2 spike protein were collected, washed in wash buffer (PBS, 1% BSA, 2 mM EDTA) and resuspended in PBS. Ninety-thousand cells per well were dispensed into

round bottom 96-well plates (Corning), and treated with 10 µg ml⁻¹ TPCK-Trypsin (Worthington Biochem) for 30 min at 37 °C. Cells were washed and incubated with 15 µg ml⁻¹ antibody across 5, 30, 60, 120 and 180 min time points at 37 °C. Cells were washed with ice-cold wash buffer, and stained with 1.5 µg ml⁻¹ Alexa Fluor647-conjugated goat anti-human IgG secondary antibody (Jackson Immunoresearch) for 30 min on ice in the dark. Cells were washed twice with cold wash buffer and analysed using a ZE5 cytometer (Bio-Rad) with acquisition chamber at 4 °C. Binding at each time point was measured as mean fluorescence intensity (MFI), normalized to the MFI at the 5 min labelling time point. Data were analysed and plotted using GraphPad Prism v. 9.0.1.

Cell-cell fusion of CHO-S cells

Cell–cell fusion between S-expressing CHO-K1 cells was performed as described by Lempp et al. ³¹. CHO-K1 cells stably expressing the prototypic SARS-CoV-2 spike protein were seeded in 96-well plates (Thermo Fisher Scientific) at 12,500 cells per well. The following day, antibody and nuclei marker Hoechst (final dilution 1:1,000) were added to cells and incubated for 24 h. Cell–cell fusion was visualized using the Cytation 5 Imager (BioTek), and an object detection protocol was used to detect nuclei and measure their size. The nuclei of fused cells (syncytia) are aggregated at the centre of the syncytia and recognized as a uniquely large object that is gated according to its size. To quantify cell–cell fusion, we report the area of objects in fused cells divided by the total area of all objects, multiplied by 100% to represent it as a percentage.

Antibody blockade of RBD binding to ACE2

ACE2 blockade ELISA was performed as described by Piccoli et al. ¹⁵. Unlabelled antibodies were serially diluted, mixed with RBD mouse Fc-tagged antigen (Sino Biological, final concentration 20 ng ml⁻¹) and incubated for 30 min at 37 °C. The mix was added for 30 min to ELISA 96-well plates (Corning) pre-coated overnight at 4 °C with 2 µg ml⁻¹ human ACE2 in PBS. Plates were washed and RBD binding was revealed using secondary goat anti-mouse IgG (Southern Biotech 1030-04). After washing,

pNPP substrate was added and plates were read at 405 nm. The percentage of inhibition was calculated as: $(1 - (\text{OD}_{\text{sample}} - \text{OD}_{\text{neg ctrl}})/(\text{OD}_{\text{pos ctrl}} - \text{OD}_{\text{neg ctrl}})) \times 100\%$, where $\text{OD}_{\text{sample}}$, $\text{OD}_{\text{neg ctrl}}$ and $\text{OD}_{\text{pos ctrl}}$ represent the optical densities of the sample, negative control and positive control, respectively.

Inhibition of spike-mediated cell–cell fusion

Cell–cell fusion inhibition assays were performed as described by McCallum et al. ⁸⁰. Vero E6 cells were seeded in 96-well plates at 15,000 cells per well in 70 μl DMEM with high glucose and 2.4% FBS (Hyclone). After 16 h at 37 °C with 8% CO₂, the cells were transfected as follows: for 10 wells, 0.57 μg plasmid SARS-CoV-2-S-D19_pcDNA3.1 was mixed with 1.68 μl X-tremeGENE HP in 30 μl OPTIMEM. After 15 min incubation, the mixture was diluted 1:10 in DMEM medium and 30 μl was added per well. Fourfold antibody serial dilutions were prepared and added to the cells, with a starting concentration of 20 $\mu\text{g ml}^{-1}$. The following day, 30 μl 5×-concentrated DRAQ5 in DMEM was added per well and incubated for 2 h at 37 °C. Nine images of each well were acquired with a Cytation 5 equipment for analysis.

S2H97 prophylactic protection in Syrian hamsters

We used a validated SARS-CoV-2 Syrian golden hamster model of infection^{81,82} to test S2H97 prophylactic efficacy. Experiments were performed in the high-containment A3 and BSL3+ facilities of the KU Leuven Rega Institute (3CAPS) under licenses AMV 30112018 SBB 219 2018 0892 and AMV 23102017 SBB 219 20170589 according to institutional guidelines.

Syrian hamsters (*Mesocricetus auratus*) were purchased from Janvier Laboratories. Hamsters were housed per two in ventilated isolator cages (IsoCage N Biocontainment System, Tecniplast) with ad libitum access to food, water, and cage enrichment (wood block). Housing conditions and experimental procedures were approved by the ethical committee of animal experimentation of KU Leuven (license P065-2020). Sample sizes of 6

hamsters was determined in order to have a significant difference of at least 1 log(viral RNA level) (effect size $d = 2.004$) between control and treatment groups, by using a two-tailed t -test with 80% power and an alpha of 0.05, calculated with G*Power 3.1 software. Six- to ten-week-old female hamsters were randomized for administration of 25 mg kg⁻¹ S2H97 antibody or 20 mg kg⁻¹ human isotype control via intraperitoneal injection. Approximately 5 h before infection, animals were anesthetized with isoflurane to allow collection of a blood sample from the jugular vein to be used for antibody quantification. Forty-eight hours after antibody injection, hamsters were infected intranasally with 1.89×10^6 TCID₅₀ SARS-CoV-2 virus in 50 µl inoculum. The challenge virus was a SARS-CoV-2 Wuhan isolate from February, 2020 (EPI_ISL_407976), passaged on Vero E6 cells. Passage 6 stock titre was determined by end-point dilution on Vero E6 cells by the Reed and Muench method⁸³, expressed as 50% tissue culture infectious dose (TCID₅₀).

Hamsters were monitored for appearance, behaviour, and weight. At day 4 post-infection, hamsters were euthanized by intraperitoneal injection of 500 µl Dolethal (200 mg ml⁻¹ sodium pentobarbital, Vétoquinol SA). Lungs were collected, homogenized via bead disruption (Precellys) in 350 µl RLT buffer (RNeasy Mini kit, Qiagen) and centrifuged (10,000 rpm, 5 min, 4 °C) to pellet cell debris. RNA was extracted using a NucleoSpin kit (Macherey-Nagel) according to manufacturer instructions. Quantitative PCR with reverse transcription (RT-qPCR) was performed on a LightCycler96 platform (Roche) using the iTaq Universal Probes One-Step RT-qPCR kit (Bio-Rad) with N2 primers and probes targeting the nucleocapsid⁸¹. Standards of SARS-CoV-2 cDNA (IDT) were used to express viral genome copies per mg tissue. To quantify infectious SARS-CoV-2 particles, end-point titrations were performed on confluent Vero E6 cells in 96-well plates. Viral titres were calculated as above, and were expressed as TCID₅₀ per mg tissue. The circulating antibody levels were measured by mesoscale bridging ELISA, using an anti-human LS mutation monoclonal antibody as a capture and anti-human CH2 monoclonal antibody as detection. Technicians performing RNA, virus, and antibody quantification were blinded to the treatment groups of processed samples. RNA and viral levels were compared between treatment and control via two-tailed Mann-

Whitney test, excluding the two treatment animals with undetectable S2H97 levels at time of viral challenge.

Blockade of binding serology competition assays

Sera blockade of antibody binding was performed as described in Piccoli et al. [15](#). In brief, human IgG1 antibodies were biotinylated using the EZ-link NHS-PEO solid phase biotinylation kit (Pierce). Each labelled antibody was tested for binding to RBD by ELISA, and a concentration for each antibody competition experiment was selected to achieve 80% maximal binding (EC_{80}). ELISA 96-well plates (Corning) were pre-coated overnight at 4 °C with 1 µg ml⁻¹ of mouse Fc-tagged RBD antigen (Sino Biological) in PBS. Unlabelled sera or plasma were serially diluted and added to ELISA plates for 30 min, followed by addition of biotinylated anti-RBD antibody at its EC_{80} concentration. After 30 min incubation, plates were washed and antibody binding was detected using alkaline phosphatase-conjugated streptavidin (Jackson ImmunoResearch). Plates were washed, pNPP substrate (Sigma-Aldrich) was added, and plates were read at 405 nm. The percentage of inhibition of antibody binding was calculated as: $(1 - (OD_{sample} - OD_{neg\ ctrl}) / (OD_{pos\ ctrl} - OD_{neg\ ctrl})) \times 100$.

Selection of VSV-SARS-CoV-2 monoclonal antibody resistance mutants

VSV-SARS-CoV-2 S chimera was used to select for SARS-CoV-2 S monoclonal antibody resistant mutants (MARMs) as previously described [1,84](#). In brief, MARMs were recovered by plaque isolation on Vero E6 cells (ATCC, CRL-1586) with the indicated monoclonal antibody in the overlay. The concentration of monoclonal antibody in the overlay was determined by neutralization assays at a MOI of 100. Escape clones were plaque-purified on Vero cells (ATCC, CCL-81) in the presence of monoclonal antibody, and plaques in agarose plugs were amplified on MA104 cells (a gift from H. B. Greenberg (Stanford School of Medicine)) with the monoclonal antibody present in the medium. Viral stocks were amplified on MA104 cells at an MOI of 0.01 in Medium 199 containing 2% FBS and 20 mM HEPES pH 7.7 (Millipore Sigma) at 34 °C. Viral supernatants were collected upon extensive cytopathic effect and clarified of cell debris by centrifugation at 1,000g for 5 min. Aliquots were

maintained at -80°C . Viral RNA was extracted from VSV-SARS-CoV-2 mutant viruses using RNeasy Mini kit (Qiagen), and S was amplified using OneStep RT–PCR Kit (Qiagen). The mutations were identified by Sanger sequencing (Genewiz). Their resistance was verified by subsequent virus infection in the presence or absence of antibody. In brief, Vero cells were seeded into 12-well plates for overnight. The virus was serially diluted using DMEM and cells were infected at 37°C for 1 h. Cells were cultured with an agarose overlay in the presence or absence of monoclonal antibody at 34°C for 2 days. Plates were scanned on a biomolecular imager and expression of eGFP is shown at 48 h after infection. The S2X58-selected mutation S494L is not shown in Fig. 3a, as its effect on RBD expression was below the deep mutational scanning computational filter.

Viral replication fitness assays

Vero E6 cells (ATCC, CRL-1586) were seeded at 1×10^6 cells per well in 6-well plates. Cells were infected with MOI of 0.02, with wild-type and four mutant VSV-SARS-CoV-2 S chimeras mixed at equal (0.20) frequencies. Following 1 h incubation, cell monolayers were washed 3 times with HBBS and cultures were incubated for 72 h in humidified incubators at 34°C . To passage the progeny viruses, virus mixture was continuously passaged 4 times in Vero E6 cells at MOI of 0.02. Cellular RNA samples from each passages were extracted using RNeasy Mini kit (QIAGEN) and subjected to next-generation sequencing as described previously to confirm the introduction and frequency of substitutions⁸⁴.

Molecular dynamics simulations

Full details of molecular dynamics workflow and analysis are available on GitHub: <https://github.com/choderalab/rbd-ab-contact-analysis>. The RBD–S309 complex was constructed from PDB 7JX3 (chains A, B and R). 7JX3 was first refined using ISOLDE⁶⁷. Refinement included adjusting several rotamers, flipping several peptide bonds, fixing several weakly resolved waters, and building in a missing four-residue-long loop. Though the N343 glycan *N*-Acetylglucosamine (NAG) was present in 7JX3, ISOLDE was used to construct a complex glycan at N343. The full glycosylation pattern

was determined from refs. ^{85,86}. The glycan structure used for N343 (FA2G2) corresponds to the most stable conformer obtained from multi microsecond molecular dynamics simulations of cumulative sampling⁸⁷. The base NAG residue in FA2G2 was aligned to the corresponding NAG stub in the RBD-S309 model and any resulting clashes were refined in ISOLDE. The same process was repeated for the RBD–S2H97 crystal structure.

The refined glycosylated RBD–S309 and RBD–S2H97 complexes were prepared for simulation using tleap from AmberTools20⁸⁸. All relevant disulfide bridges and covalent connections in glycan structures were specified. The glycosylated proteins were parameterized with the Amber ff14SB⁸⁹ and GLYCAM_06j-1⁹⁰ force fields. The systems were solvated using the TIP3P rigid water model⁹¹ in a truncated octahedral box with 2.2 nm solvent padding on all sides. The shape and size of the solvent box were chosen to prevent the protein complex from interacting with its periodic image. The solvated systems were then neutralized with 0.15 M NaCl using the Li/Merz ion parameters of monovalent ions for the TIP3P water model (12-6 normal usage set)⁹². Virtual bonds were added across chains that should be imaged together to aid the post-processing of trajectories.

Each system was energy-minimized with an energy tolerance of 10 kJ mol⁻¹ and equilibrated five times independently using the OpenMMTools 0.20.0 (<https://github.com/choderalab/openmmtools>) BAOAB Langevin integrator⁹³ for 20 ns in the NPT ($P = 1$ atm, $T = 310$ K) ensemble with a time step of 4.0 fs, a collision rate of 1.0 ps⁻¹, and a relative constraint tolerance of 1×10^{-5} . Hydrogen atom masses were set to 4.0 amu by transferring mass from connected heavy atoms, bonds to hydrogen were constrained, and centre of mass motion was not removed. Pressure was controlled by a molecular-scaling Monte Carlo barostat with an update interval of 25 steps. Non-bonded interactions were treated with the Particle Mesh Ewald method⁹⁴ using a real-space cut-off of 1.0 nm and the OpenMM default relative error tolerance of 0.0005, with grid spacing selected automatically. The simulations were subsequently packaged to seed for production simulation on Folding@home^{95,96}. Default parameters were used unless noted otherwise.

The equilibrated structures (five per complex) were used to initiate parallel distributed molecular dynamics simulations on Folding@home^{95,96}.

Simulations were run with OpenMM 7.4.2 (compiled into Folding@home core22 0.0.13). Production simulations used the same Langevin integrator as the NPT equilibration described above. A total of 5,000 and 4,985 independent molecular dynamics simulations were generated on Folding@home for RBD–S309 and RBD–S2H97, respectively.

Conformational snapshots (frames) were stored at an interval of 1 ns per frame for subsequent analysis. The final datasets contained 1.1 ms and 623.7 μ s of aggregate simulation time for RBD–S309 and RBD–S2H97, respectively. This trajectory dataset (without solvent) are available at the MolSSI COVID-19 molecular structure and therapeutics hub (<https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-monoclonal-antibody-s309> and <https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-monoclonal-antibody-s2h97>).

The first 100 ns of each trajectory was discarded (to allow relaxation away from the crystal structure), yielding total simulation times of 644.3 and 262.9 μ s used for analysis of RBD–S309 and RBD–S2H97 systems, respectively. All trajectories had solute structures aligned to their first frame and centred using MDTraj⁹⁷. Residues were considered to be at the interface if they were within 10 Å of any antibody Fab / RBD residue (with the exception of RBD N343 glycans, where all glycan residues were considered). The minimum distance of heavy atoms between every pair of interface residues was computed for every frame (1 ns) using MDAnalysis^{98,99}. A close contact was counted if the minimum distance between a residue pair was below 3.5 Å (if one of the residues was hydrophobic, a 4.5 Å cut-off was used). The contribution of each RBD residue to close contacts was calculated as a percentage by summation of the number of close contacts for a particular RBD residue and normalizing by the total number of close contact interactions over all frames of each simulation.

Materials availability

The SARS-CoV-2 RBD mutant libraries (#1000000172) and unmutated parental plasmid (#166782) are available on Addgene. Other materials generated in this study will be made available on request and may require a material transfer agreement.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Interactive escape maps and structural visualizations can be found at https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Vir_mAbs/. Raw Illumina sequencing data from deep mutational scanning experiments are available on NCBI Sequence Read Archive BioSample under accessions [SAMN18315604](#) (SARS-CoV-2 mutant selection data) and [SAMN18316011](#) (sarbecovirus RBD selection data). PacBio sequencing data used to link N16 barcodes to sarbecovirus RBD variant are available on NCBI SRA BioSample under accession [SAMN18316101](#). A complete table of deep mutational scanning antibody escape fractions is provided on GitHub: https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs/blob/main/results/supp_data/all_antibodies_raw_data.csv. This table includes both antibodies first described in this study (Fig. 1b, c), and all other antibody selections that were re-processed to generate Fig. 4a. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code [7R6W](#) for RBD-S2X35-S309, [7M7W](#) for RBD-S2H97-S2X259 and [7R6X](#) for RBD-S2E12-S304-S309. Cryo-electron microscopy structure data and models are available from the Electron Microscopy Data Bank (EMDB) under accession codes [EMD-24300](#) for S/S2D106, EMDB [EMD-24299](#) and PDB [7R7N](#) for the S/S2D106 local refinement, and EMDB [EMD-24301](#) for S/S2H97. The raw and processed molecular dynamics trajectory data are available at the MolSSI COVID-19 Molecular Structure and Therapeutics Hub: <https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-monoclonal-antibody-s309> and

<https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-monoclonal-antibody-s2h97>. All other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. [Source data](#) are provided with this paper.

Code availability

A repository containing all code, analysis and summary notebooks for the analysis of the SARS-CoV-2 deep mutational scanning escape selections is available on GitHub at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs. A repository containing code and analysis of the sarbecovirus RBD library binding experiments is available on GitHub at https://github.com/jbloomlab/SARS-CoV_RBD_MAP. A repository containing code and analysis of molecular dynamics simulations is available on GitHub at <https://github.com/choderalab/rbd-ab-contact-analysis>.

References

1. 1.

Liu, Z. et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **29**, 477–488.e4 (2021).

2. 2.

Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).

3. 3.

Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57.e9 (2021).

4. 4.

Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).

5. 5.

Rappazzo, C. G. et al. Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. *Science* **371**, 823–829 (2021).

6. 6.

Menachery, V. D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).

7. 7.

Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).

8. 8.

Tortorici, M. A. et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* **370**, 950–957 (2020).

9. 9.

Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

10. 10.

Zost, S. J. et al. Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**, 443–449 (2020).

11. 11.

Hassan, A. O. et al. A SARS-CoV-2 infection model in mice demonstrates protection by neutralizing antibodies. *Cell* **182**, 744–753.e4 (2020).

12. 12.

Winkler, E. S. et al. Human neutralizing antibodies against SARS-CoV-2 require intact Fc effector functions for optimal therapeutic protection. *Cell* **184**, 1804–1820.e16 (2021).

13. 13.

Schäfer, A. et al. Antibody potency, effector function, and combinations in protection and therapy for SARS-CoV-2 infection in vivo. *J. Exp. Med.* **218**, e20201993 (2021).

14. 14.

Corti, D., Purcell, L. A., Snell, G. & Veesler, D. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* **184**, 3086–3108 (2021).

15. 15.

Piccoli, L. et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).

16. 16.

Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476.e6 (2021).

17. 17.

Chen, R. E. et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).

18. 18.

Wang, P. et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).

19. 19.

Thomson, E. C. et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187.e20 (2021).

20. 20.

Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).

21. 21.

Starr, T. N., Greaney, A. J., Dingens, A. S. & Bloom, J. D. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep. Med.* **2**, 100255 (2021).

22. 22.

Cathcart, A. L. et al. The dual function monoclonal antibodies VIR-7831 and VIR-7832 demonstrate potent in vitro and in vivo activity against SARS-CoV-2. Preprint at <https://doi.org/10.1101/2021.03.09.434607> (2021).

23. 23.

McCallum, M. et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* <https://doi.org/10.1126/science.abi7994> (2021).

24. 24.

Gupta, A. et al. Early Covid-19 treatment with SARS-CoV-2 neutralizing antibody sotrovimab. Preprint at <https://doi.org/10.1101/2021.05.27.21257096> (2021).

25. 25.

Dingens, A. S., Arenz, D., Weight, H., Overbaugh, J. & Bloom, J. D. An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* **50**, 520–532.e3 (2019).

26. 26.

Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).

27. 27.

Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).

28. 28.

Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).

29. 29.

Zhang, W. et al. Emergence of a novel SARS-CoV-2 variant in southern California. *J. Am. Med. Assoc.* **325**, 1324–1326 (2021).

30. 30.

Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039.e15 (2019).

31. 31.

Lempp, F. A. et al. Membrane lectins enhance SARS-CoV-2 infection and influence the neutralizing activity of different classes of antibodies. Preprint at <https://doi.org/10.1101/2021.04.03.438258> (2021).

32. 32.

Pinto, D. et al. A human antibody that broadly neutralizes betacoronaviruses protects against SARS-CoV-2 by blocking the fusion machinery. Preprint at <https://doi.org/10.1101/2021.05.09.442808> (2021).

33. 33.

Walls, A. C. et al. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).

34. 34.

Cohen, A. A. et al. Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021).

35. 35.

Saunders, K. O. et al. Neutralizing antibody vaccine for pandemic and pre-emergent coronaviruses. *Nature* **594**, 553–559 (2021).

36. 36.

Dong, J. et al. Genetic and structural basis for recognition of SARS-CoV-2 spike protein by a two-antibody cocktail. Preprint at <https://doi.org/10.1101/2021.01.27.428529> (2021).

37. 37.

Tortorici, M. A. et al. Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* <https://doi.org/10.1038/s41586-021-03817-4> (2021).

38. 38.

Sauer, M. M. et al. Structural basis for broad coronavirus neutralization. *Nat. Struct. Mol. Biol.* **28**, 478–486 (2021).

39. 39.

Kistler, K. E. & Bedford, T. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. *eLife* **10**, e64509 (2021).

40. 40.

Eguia, R. T. et al. A human coronavirus evolves antigenically to escape antibody immunity. *PLoS Pathog.* **17**, e1009453 (2021).

41. 41.

Walls, A. C. et al. Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines. Preprint at <https://doi.org/10.1101/2021.03.15.435528> (2021).

42. 42.

Zhou, P. et al. A protective broadly cross-reactive human antibody defines a conserved site of vulnerability on beta-coronavirus spikes. Preprint at <https://doi.org/10.1101/2021.03.30.437769> (2021).

43. 43.

Crawford, K. H. D. et al. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* **12**, 513 (2020).

44. 44.

Xie, X. et al. A nanoluciferase SARS-CoV-2 for rapid neutralization testing and screening of anti-infective drugs for COVID-19. *Nat. Commun.* **11**, 5214 (2020).

45. 45.

Takada, A. et al. A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94**, 14764–14769 (1997).

46. 46.

Riblett, A. M. et al. A haploid genetic screen identifies heparan sulfate proteoglycans supporting rift valley fever virus infection. *J. Virol.* **90**, 1414–1423 (2015).

47. 47.

Giroglou, T. et al. Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein. *J. Virol.* **78**, 9007–9015 (2004).

48. 48.

Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).

49. 49.

Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e6 (2020).

50. 50.

Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis. *Proc. Natl Acad. Sci. USA* **115**, E7550–E7558 (2018).

51. 51.

Hilton, S. K. et al. *dms-view*: Interactive visualization tool for deep mutational scanning data. *J. Open Source Softw.* **5**, 2353 (2020).

52. 52.

Song, H.-D. et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl Acad. Sci. USA* **102**, 2430–2435 (2005).

53. 53.

Tong, S. et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* **15**, 482–485 (2009).

54. 54.

Zhou, H. et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**, 2196–2203.e3 (2020).

55. 55.

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

56. 56.

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

57. 57.

Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).

58. 58.

Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

59. 59.

Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).

60. 60.

Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

61. 61.

Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).

62. 62.

Hansen, J. et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**, 1010–1014 (2020).

63. 63.

Shi, R. et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).

64. 64.

Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).

65. 65.

McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).

66. 66.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

67. 67.

Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).

68. 68.

Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).

69. 69.

Hsieh, C.-L. et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).

70. 70.

Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

71. 71.

Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).

72. 72.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure

determination. *Nat. Methods* **14**, 290–296 (2017).

73. 73.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

74. 74.

Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

75. 75.

Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

76. 76.

Ramlau, K., Palmer, C. M., Nakane, T. & Aylett, C. H. S. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. *J. Struct. Biol.* **211**, 107545 (2020).

77. 77.

Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

78. 78.

Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078 (2020).

79. 79.

Wang, R. Y.-R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).

80. 80.

McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e16 (2021).

81. 81.

Boudewijns, R. et al. STAT2 signaling restricts viral dissemination but drives severe pneumonia in SARS-CoV-2 infected hamsters. *Nat. Commun.* **11**, 5838 (2020).

82. 82.

Sanchez-Felipe, L. et al. A single-dose live-attenuated YF17D-vectored SARS-CoV-2 vaccine candidate. *Nature* **590**, 320–325 (2021).

83. 83.

Reed, L. J. & Muench, H. A simple method of estimating fifty percent endpoints. *Am. J. Epidemiol.* **27**, 493–497 (1938).

84. 84.

Case, J. B. et al. Neutralizing antibody and soluble ACE2 inhibition of a replication-competent VSV-SARS-CoV-2 and a clinical isolate of SARS-CoV-2. *Cell Host Microbe* **28**, 475–485.e5 (2020).

85. 85.

Shajahan, A. et al. Comprehensive characterization of *N*- and *O*-glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. *Glycobiology* **31**, 410–424 (2020).

86. 86.

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330–333 (2020).

87. 87.

Harbison, A. M., Brosnan, L. P., Fenlon, K. & Fadda, E. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94–103 (2019).

88. 88.

Case, D.A. et al. *Amber 2020* (Univ. of California, 2020).

89. 89.

Maier, J. A. et al. Ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

90. 90.

Kirschner, K. N. et al. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 (2008).

91. 91.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

92. 92.

Li, P., Song, L. F. & Merz, K. M., Jr Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* **11**, 1645–1657 (2015).

93. 93.

Leimkuhler, B. & Matthews, C. Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys.* **138**, 174102 (2013).

94. 94.

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**, 10089–10092 (1993).

95. 95.

Shirts, M. & Pande, V. S. COMPUTING: Screen savers of the world unite! *Science* **290**, 1903–1904 (2000).

96. 96.

Zimmerman, M. I. et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **13**, 651–659 (2021).

97. 97.

McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

98. 98.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

99. 99.

Gowers, R. et al. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. in *Proc. 15th Python in*

Science Conference (SciPy, 2016).

Acknowledgements

We thank I. Hoffman for assistance in refinement of crystal structures, G. R. Bowman, J. Coffland and P. K. Eastman for developing and maintaining the Folding@home infrastructure, Amazon Web Services for Folding@home infrastructure support, the Folding@home volunteers who contributed their computational resources to this project (FAH Project 17336-17340), R. P. Wiewiora and S. Singh for their assistance with Folding@home, and A. M. Harbison and E. Fadda for assistance in glycan modeling. This work was supported by the Fred Hutch Flow Cytometry and Genomics facilities, the Fred Hutch Scientific Computing group supported by ORIP grant S10OD028685, and the University of Washington Arnold and Mabel Beckman Cryo-EM Center. This work was supported by the NIH/NIAID (R01AI127893 and R01AI141707 to J.D.B., DP1AI158186 and HHSN272201700059C to D.V. and T32AI083203 to A.J.G.), the NIH/NIGMS (R01GM120553 to D.V., and grant R01GM121505 and R01GM132386 to J.D.C.), the NIH/NCI (P30CA008748 to J.D.C.), the National Science Foundation (NSF CHI-1904822 to J.D.C.), the Damon Runyon Cancer Research Foundation (T.N.S.), the Gates Foundation (INV-004949 to J.D.B.), a Pew Biomedical Scholars Award (D.V.), Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (J.D.B. and D.V.), the Wellcome Trust (209407/Z/17/Z to T.I.C.), Fast Grants (D.V.), Bayer (W.G.G.) and the Molecular Sciences Software Institute (I.Z.). J.D.B. is an Investigator of the Howard Hughes Medical Institute. The Molecular Biology Consortium beamline 4.2.2 of the Advanced Light Source, a US DOE Office of Science User Facility under contract no. DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the NIH/NIGMS (P30GM124169-01). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH/NIGMS (P30GM133894).

Author information

Author notes

1. These authors contributed equally: Tyler N. Starr, Nadine Czudnochowski, Zhuoming Liu

Affiliations

1. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Tyler N. Starr, Amin Addetia, Allison J. Grianey, Adam S. Dingens & Jesse D. Bloom

2. Vir Biotechnology, San Francisco, CA, USA

Nadine Czudnochowski, Patrick Hernandez, Jason A. Wojcechowskyj, Laura E. Rosen, Jiayi Zhou, Martin Montiel-Ruiz, Hannah Kaiser, Josh R. Dillen, Heather Tucker, Michael P. Housley, Julia di Julio, Maria Agostini, Marcel Meury, Exequiel Dellota Jr, Colin Havenar-Daughton, Amilio Telenti, Florian A. Lempp, Christy M. Hebner, Herbert W. Virgin & Gyorgy Snell

3. Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO, USA

Zhuoming Liu, Spencer Stumpf & Sean P. J. Whelan

4. Humabs BioMed SA, a subsidiary of Vir Biotechnology, Bellinzona, Switzerland

Fabrizia Zatta, Dora Pinto, Martina Beltramello, Roberta Marzi, Anna De Marco, Jessica Bassi, Chiara Silacci-Fregni, Gloria Lombardo, Nicole Sprugasci, Katja Culap, Stefano Jaconi, Elisabetta Cameroni, Luca Piccoli, Fabio Benigni, Matteo S. Pizzuto & Davide Corti

5. Department of Biochemistry, University of Washington, Seattle, WA, USA

Young-Jun Park, John E. Bowen, M. Alejandra Tortorici, Alexandra C. Walls & David Veesler

6. Department of Genome Sciences, University of Washington, Seattle, WA, USA

Allison J. Greaney & Jesse D. Bloom

7. Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

William G. Glass, Ivy Zhang & John D. Chodera

8. Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA

Ivy Zhang

9. Rega Institute for Medical Research, Laboratory of Virology and Chemotherapy, KU Leuven, Leuven, Belgium

Rana Abdelnabi, Shi-Yan Caroline Foo & Johan Neyts

10. Cambridge Institute for Medical Research, Department of Haematology, University of Cambridge, Cambridge, UK

Tristan I. Croll

11. Molecular Biology Consortium, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Jay C. Nix

12. Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA

Herbert W. Virgin

13. Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX, USA

Herbert W. Virgin

14. Howard Hughes Medical Institute, Seattle, WA, USA

Jesse D. Bloom

Contributions

Conceived research and designed study: T.N.S., N.C., H.W.V., D.C., J.D.B. and G.S. Antibody discovery: F.Z., D.P., M.B., R.M., A.D.M., E.C., M.S.P. and D.C. Expression and purification of proteins: N.C., P.H., G.L., N.S., J.E.B., A.C.W., K.C., S.J. and M.M. Antibody functional experiments: Y.-J.P., Z.L., F.Z., D.P., M.B., R.M., J.E.B., M.A.T., A.C.W., J.A.W., A.D.M., L.E.R., J.Z., M.M.-R., H.K., J.R.D., H.T., J.B., C.S.-F., M.P.H., M.A., E.D.Jr, S.S., C.H.-D., L.P., F.B., F.A.L. and S.P.J.W. Deep mutational scanning experiments and analysis: T.N.S., A.A., A.J.G. and A.S.D. Hamster model: R.A., S.-Y.C.F., F.B. and J.N. Bioinformatics analysis: J.d.I. and A.T. Structure determination: N.C., Y.-J.P., P.H., J.E.B., T.I.C., J.C.N., D.V. and G.S. Molecular dynamics simulation and analysis: W.G.G., I.Z. and J.D.C. Supervision: M.S.P., J.D.C., C.M.H., S.P.J.W., D.V., D.C., J.D.B. and G.S. Writing of the initial draft: T.N.S., N.C., D.C., J.D.B. and G.S. Editing of the final version: all authors.

Corresponding authors

Correspondence to [Davide Corti](#) or [Jesse D. Bloom](#) or [Gyorgy Snell](#).

Ethics declarations

Competing interests

N.C., F.Z., D.P., M.B., P.H., R.M., J.A.W., A.D.M., L.E.R., J.Z., M.M.-R., H.K., J.R.D., H.T., J.B., C.S.-F., M.P.H., J.d.I., G.L., M.A., N.S., K.C., S.J., M.M., E.D.Jr, E.C., C.H.-D., L.P., F.B., A.T., F.A.L., M.S.P., C.M.H., H.W.V., D.C. and G.S. are or were employees of Vir Biotechnology and may hold shares in Vir Biotechnology. D.C. is currently listed as an inventor on multiple patent applications, which disclose the subject matter described in this manuscript. After the submission of the initial version of this study, J.D.B. began consulting for Moderna on viral evolution and epidemiology. J.D.B. has the potential to receive a share of IP revenue as an inventor on a Fred Hutchinson Cancer Research Center-optioned technology/patent (application WO2020006494) related to deep mutational scanning of viral proteins. H.W.V. is a founder of PierianDx and Casma Therapeutics. Neither company provided funding for this work nor is performing related work. J.C.N., T.I.C. and D.V. are consultants for Vir Biotechnology Inc. The Veesler laboratory has received a sponsored research agreement from Vir Biotechnology Inc. J.D.C. is a current member of the Scientific Advisory Boards of OpenEye Scientific Software, Interline Therapeutics, and Redesign Science. The Chodera laboratory receives or has received funding from the National Institute of Health, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, Vir Biotechnology, XtalPi, Interline Therapeutics, and the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at <http://choderelab.org/funding>. The other authors declare no competing interests.

Additional information

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Antibody neutralization and binding data.

a, Neutralization of authentic SARS-CoV-2 (SARS-CoV-2-Nluc) by 14 antibodies. Shown are representative live virus neutralization plots, measured with entry into Vero E6 cells. Symbols are means \pm s.d. of technical triplicates. Dashed lines indicate IC₅₀ and IC₉₀ values. All antibodies were measured at each concentration point in the series, with hidden points due to overplotting reflecting overlap at the upper and lower neutralization limits. **b**, Correlation in antibody neutralization IC₅₀ as determined in spike-pseudotyped VSV particles ($n = 3$ to 8) versus authentic SARS-CoV-2 ($n = 3$). **c**, Representative SPR sensorgrams of Fab fragments of the six newly described antibodies binding to the SARS-CoV-2 RBD (Fig. 1a gives average values for $n = 2$ –4 measurements). White and grey stripes indicate association and dissociation phases, respectively. Data shown in blue and fit to binding model shown in dashed gray line. Binding affinities for previously described antibodies shown in Fig. 1a are consistent with measurements from Piccoli et al. (S304, S309, S2X35, S2H13, S2H14)¹⁵ and Tortorici et al. (S2E12)⁸. **d**, Identifiers and spike genotypes of SARS-CoV-2 variants tested in neutralization assays in Figs. 2d, 3b.

Extended Data Fig. 2 Deep mutational scanning to map mutations that escape antibody binding.

a, Scheme of the deep mutational scanning assay. Conformationally intact RBD is expressed on the surface of yeast, where RBD expression and antibody binding is detectable via fluorescent labelling. We previously constructed mutant libraries containing virtually all of the 3,819 possible amino acid mutations in the SARS-CoV-2 RBD²⁶ and sorted the library to eliminate mutations that destabilize the RBD or strongly reduce ACE2-

binding affinity³. We incubate the library with a sub-saturating antibody concentration and use fluorescence-activated cell sorting (FACS) to isolate yeast cells expressing RBD mutants with reduced antibody binding. Deep sequencing quantifies mutant frequencies before and after FACS selection, enabling calculation of the “escape fraction” of each amino acid mutation, which reflects the fraction of cells carrying that mutation that fall into the antibody-escape bin. Mutation escape fractions are represented in logo plots, where the height of a letter reflects the extent of escape from antibody binding. **b**, Representative selection gates, after gating for single cells expressing RBD as in Greaney et al.³. Yeast expressing the SARS-CoV-2 RBD (top panels) are labelled at 1x, 0.01x and no antibody to guide selection gates. Mutant RBDs that reduce binding (green, gate drawn to capture 0.01x wild-type control) are sorted and sequenced for calculation of mutant escape fractions. This same gate was used to quantify escape within libraries of yeast expressing all sarbecovirus RBD homologues. For several antibodies, we also selected the sarbecovirus RBD library with a more stringent “full escape” gate (blue, gate drawn to capture 0 ng ml⁻¹ wild-type control). **c**, Fraction of library cells falling into escape bins for each antibody selection. **d**, Line plots showing total escape at all RBD sites for each antibody. Sites of strong escape illustrated in logo plots in Fig. [1b,c](#) shown with pink indicators. **e, f**, Correlation in per-mutation (**e**) and per-site (**f**, sum of per-mutation) escape fractions for duplicate libraries that were independently generated and assayed. *N*, number of mutations (**e**) or sites (**f**) in the correlation.

[Extended Data Fig. 3 Antibody escapability from deep mutational scanning measurements and in natural SARS-CoV-2 mutants.](#)

a, To calculate antibody escapability (Fig. [1b,c](#)), mutation escape fractions were weighted by their deleterious consequences for ACE2 binding or RBD expression. Top plots show the weighting factor (*y*-axis) for mutation effects on ACE2 binding (left) and RBD expression (right). This weight factor was multiplied by the mutation escape fraction in the summation to calculate antibody escapability as described in the Methods. Histograms show the distribution of mutation effects on ACE2 binding (left) and RBD

expression (right) for all mutations that pass our computational filtering steps (bottom), and mutations that are found with at least 20 sequence counts on GISAID (middle). **b**, Correlation in antibody relative epitope size (top) and escapability (bottom) calculated from independent deep mutational scanning replicates, compared to the averaged replicates shown in Fig. [1b](#),[c](#). R^2 , squared Pearson correlation coefficient. **c**, Scatter plots illustrate the degree to which a mutation escapes antibody binding (escape fraction, y-axis) and its frequency among 1,190,241 high-quality human-derived SARS-CoV-2 sequences present on GISAID as of May 2, 2021. Large escape mutations (>5 x global median escape fraction) for each antibody with non-zero mutant frequencies are labelled. Plot labels report the sum of mutant frequencies for all labelled mutations, corresponding to the natural SARS-CoV-2 mutant escape frequency for antibodies shown in Fig. [4d](#),[g](#).

Extended Data Fig. 4 Breadth of antibody binding across sarbecoviruses.

a, Phylogenetic relationship of sarbecovirus RBDs inferred from aligned nucleotide sequences, with the four sarbecovirus clades labelled in separate colours used throughout the text. Node support values are bootstrap support values. **b**, Breadth of sarbecovirus binding by each antibody to a panel of yeast-displayed sarbecovirus RBDs. Data as in Fig. [1d](#), with the addition of secondary “full escape” selection data for S2H97, S2H13, and S2H14 (0 ng ml⁻¹ wild-type control as shown in Extended Data Fig. [2b](#),[c](#)), enabling differentiation of RBDs with intermediate binding (for example, S2H97/RsSHC014) versus complete loss of binding. Escape fractions are calculated as the mean of replicate barcoded genotypes internal to the library. Median number of barcodes per RBD is 249, with a range of 104 to 566. The median SEM across escape fraction measurements is 0.019, with a range of 0.00005 to 0.038 across all RBD/antibody pairs. **c**, Flow cytometry detection of antibody binding to isogenic yeast-displayed RBD variants. **d**, Flow cytometry detection of antibody binding to mammalian-surface displayed spikes. **e**, ELISA binding of antibody to purified RBD proteins. **f**, SPR measurement of binding of cross-reactive antibodies (Fab) and human ACE2 to select sarbecovirus RBDs. NB, no binding; NT, not tested. Values from single replicates. **g**, S2H97 neutralization of VSV pseudotyped with

select sarbecovirus spikes, with entry measured in ACE2-transduced BHK-21 cells. Curves are representative of two independent experiments. Points represent means, error bars standard deviation from three technical replicates, and IC₅₀ geometric mean of experiments. IC₅₀ values are not comparable to other experiments on Vero E6 cells (for example, Fig. 2c) due to ACE2 overexpression and its impact on S2H97 neutralization. **h**, Alignment of germline-reverted and mature S2H97 heavy- (top) and light-chain (bottom) amino acid sequences. CDR sequences shown in grey box. Heat map overlay indicates the predicted energetic contribution of antibody paratope residues from the crystal structure. **i**, Binding of germline-reverted and mature S2H97 to select sarbecovirus RBDs as measured by SPR. **j**, Neutralization experiment of select sarbecoviruses by S2E12 (spike-pseudotyped VSV on 293T-ACE2 cells). Details as in Fig. 3c. **k**, Alignment of germline-reverted and mature S2E12. Details as in h. **l**, Binding of germline-reverted and mature S2E12 to select sarbecovirus RBDs as measured by SPR.

Extended Data Fig. 5 Structures and epitopes of Fab:RBD complexes.

a, Surfaces targeted by broadly binding RBD antibodies. RBD surface is coloured by site variability across sarbecoviruses. ACE2 key motifs shown in transparent red cartoon. Antibody variable domains shown as cartoon, with darker shade indicating the heavy chain. **b**, **c**, Integrative features of the S309 (**b**) and S2X35 (**c**) structural epitopes. Details as in Figs. 3g, h, 2b. Variants in the S309 epitope were tested with VIR-7831 with the exception of E340D (marked with *), which was tested with S309. **d–h**, Zoomed in view of the RBD bound to S309 (**d**), S2X35 (**e**), S2H97 (**f**), S2E12 (**g**), and S2D106 (**h**), with important contact and escape residues labelled. RBD residues coloured by total site escape [scale bar, right of (**d**)]. **i**, **j**, Representative electron micrograph and 2D class averages of SARS-CoV-2 S in complex with the S2H97 Fab embedded in vitreous ice. Scale bar: 400 Å. Micrographs representative of 3138 micrographs. **k**, Gold-standard Fourier shell correlation curve for the S2H97-bound SARS-CoV-2 S trimer reconstruction. The 0.143 cut-off is indicated by a horizontal dashed line. **l**, Local resolution map calculated using cryoSPARC for the whole reconstruction with two orthogonal orientations. **m**, **n**, Representative

electron micrograph and 2D class averages of SARS-CoV-2 S in complex with the S2D106 Fab embedded in vitreous ice. Scale bar: 400 Å.

Micrographs representative of 2166 micrographs **o**, Gold-standard Fourier shell correlation curves for the S2D106-bound SARS-CoV-2 S trimer (black line) and locally refined RBD/S2D106 variable domains (grey line). The 0.143 cut-off is indicated by a horizontal dashed line. **p**, Local resolution map calculated using cryoSPARC for the whole reconstruction and the locally refined RBD/S2D106 variable domain region.

Extended Data Fig. 6 Mechanism of action of S2H97 neutralization and protection.

a, Quaternary context of the S2H97 epitope. Left, S2H97-bound RBD, with RBD sites coloured by S2H97 escape (scale bar, bottom). Right, RBD in the same angle as left, in the closed spike trimer. **b**, CryoEM structure of S2H97 Fabs (green surfaces) bound to SARS-CoV-2 S indicating the extensive opening of the RBD (yellow surface) necessary to access the S2H97 epitope. Closed RBD (light purple surface, PDB 7K43) and site II Fab S2A4 bound open RBD (grey surface, PDB 7JVC) are shown for comparison. Spike protomers are shown in yellow, blue, and pink. **c**, Antibody-mediated S1 shedding from cell-surface expressed SARS-CoV-2 S as determined by flow cytometry. **d**, Cell-cell fusion of CHO cells expressing SARS-CoV-2 S (CHO-S) incubated with variable concentrations of antibody. **e**, Antibody competition with RBD-ACE2 binding determined by ELISA. Points represent mean of technical duplicates. **f**, S2H97 neutralization of SARS-CoV-2 S pseudotyped VSV on ACE2-overexpressing cells (293T-ACE2) compared to Vero E6 cells. Points reflect mean and error bars reflect standard deviation from triplicate measurements. Curves are representative of two biological replicates. **g**, Antibody inhibition of cell-to-cell fusion of Vero E6 cells transfected with SARS-CoV-2 S. **h**, Influence of circulating S2H97 level on prophylactic efficacy in Syrian hamsters. Infectious virus titres (right y-axis, triangles) and RNA levels (left y-axis, circles) reflect the data represented in Fig. [2f](#), measured in hamsters four days after SARS-CoV-2 challenge in animals prophylactically dosed with 25 mg/kg S2H97 (magenta symbols) or isotype control (white symbols). The levels of circulating S2H97 (D0, before infection, $\mu\text{g ml}^{-1}$) are shown on the x-axis (LLOQ, lower level of

quantification). ** $P = 0.0048$ (virus titre) and $P = 0.0048$ (RNA) vs control isotype, two-sided Mann–Whitney test (the 2 animals shown with no detectable serum antibody were excluded from the comparison).

Extended Data Fig. 7 Escapability and the relationships among antibody properties.

a, Additional spike-VSV viral escape selections, as in Fig. 3a, and an illustration of the authentic SARS-CoV-2 escape data for VIR-7832 (derived from S309) reported in Cathcart et al. 22. **b**, Correlation between the number of unique mutations selected across viral escape selection experiments and antibody escapability as tabulated in Fig. 1b,c, plus S2X259³⁷. **c**, Projected epitope space from Fig. 4a annotated by antibody properties as in Fig. 4b–d. For each property, antibodies are coloured such that purple reflects the most desirable antibody (scale bar, right; N.D., not determined): narrowest functional epitope, tightest binding affinity (K_D , log10 scale), lowest escapability. **d**, Pairwise scatter plots between all antibody properties discussed in the main text. Select scatter plots from this panel are shown in Figs. 4e–g. Details of each property described in Methods. All axes are oriented such that moving up on the y-axis and right on the x-axis corresponds to moving in the “preferred” direction for an antibody property (lower neutralization IC₅₀, lower K_D , higher breadth, narrower epitope size, lower escapability, lower total frequency of SARS-CoV-2 escape mutants among sequences on GISAID).

Extended Data Table 1 Characteristics of the antibodies described in this study

Extended Data Table 2 Crystallographic data collection and refinement statistics

Extended Data Table 3 CryoEM data collection, refinement, and validation statistics

Supplementary information

Reporting Summary

Source data

[Source Data Fig. 2](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Starr, T.N., Czudnochowski, N., Liu, Z. *et al.* SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021). <https://doi.org/10.1038/s41586-021-03807-6>

- Received: 29 March 2021
- Accepted: 06 July 2021
- Published: 14 July 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03807-6>

Further reading

- [This âsuper antibodyâ for COVID fights off multiple coronaviruses](#)
 - Diana Kwon

Nature (2021)

This ‘super antibody’ for COVID fights off multiple coronaviruses

- Diana Kwon

News 14 Jul 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03807-6>

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

- Article
- [Published: 19 July 2021](#)

Broad sarbecovirus neutralization by a human monoclonal antibody

- [M. Alejandra Tortorici](#) ORCID: orcid.org/0000-0002-2260-2577^{1,2 na1},
- [Nadine Czudnochowski](#)^{3 na1},
- [Tyler N. Starr](#) ORCID: orcid.org/0000-0001-6713-6904^{4 na1},
- [Roberta Marzi](#) ORCID: orcid.org/0000-0001-6025-6735^{5 na1},
- [Alexandra C. Walls](#) ORCID: orcid.org/0000-0002-9636-8330¹,
- [Fabrizia Zatta](#)⁵,
- [John E. Bowen](#) ORCID: orcid.org/0000-0003-3590-9727¹,
- [Stefano Jaconi](#) ORCID: orcid.org/0000-0001-7527-4434⁵,
- [Julia Di Julio](#)³,
- [Zhaoqian Wang](#)¹,
- [Anna De Marco](#)⁵,
- [Samantha K. Zepeda](#)¹,
- [Dora Pinto](#)⁵,
- [Zhuoming Liu](#) ORCID: orcid.org/0000-0001-8198-0976⁶,
- [Martina Beltramello](#)⁵,
- [Istvan Bartha](#)⁵,
- [Michael P. Housley](#)³,
- [Florian A. Lempp](#) ORCID: orcid.org/0000-0001-6103-8078³,
- [Laura E. Rosen](#) ORCID: orcid.org/0000-0002-8030-0219³,
- [Exequiel Dellota Jr](#)³,
- [Hannah Kaiser](#) ORCID: orcid.org/0000-0002-3991-7401³,
- [Martin Montiel-Ruiz](#) ORCID: orcid.org/0000-0001-6200-9578³,
- [Jiayi Zhou](#) ORCID: orcid.org/0000-0002-4231-3422³,
- [Amin Addetia](#)⁴,
- [Barbara Guarino](#)³,
- [Katja Culap](#)⁵,
- [Nicole Sprugasci](#)⁵,
- [Christian Saliba](#)⁵,
- [Eneida Vetti](#) ORCID: orcid.org/0000-0002-0216-6704⁵,

- [Isabella Giacchetto-Sasselli⁵](#),
- [Chiara Silacci Fregni⁵](#),
- [Rana Abdelnabi⁷](#),
- [Shi-Yan Caroline Foo ORCID: orcid.org/0000-0002-6380-4917⁷](#),
- [Colin Havenar-Daughton ORCID: orcid.org/0000-0002-2880-3927³](#),
- [Michael A. Schmid ORCID: orcid.org/0000-0002-1137-9322⁵](#),
- [Fabio Benigni⁵](#),
- [Elisabetta Cameroni⁵](#),
- [Johan Neyts ORCID: orcid.org/0000-0002-0033-7514⁷](#),
- [Amalio Telenti ORCID: orcid.org/0000-0001-6290-7677³](#),
- [Herbert W. Virgin ORCID: orcid.org/0000-0001-8580-7628³](#),
- [Sean P. J. Whelan ORCID: orcid.org/0000-0003-1564-8590⁶](#),
- [Gyorgy Snell ORCID: orcid.org/0000-0003-1475-659X³](#),
- [Jesse D. Bloom ORCID: orcid.org/0000-0003-1267-3408^{4,8}](#),
- [Davide Corti ORCID: orcid.org/0000-0002-5797-1364⁵](#),
- [David Veesler ORCID: orcid.org/0000-0002-6019-8675¹](#) &
- [Matteo Samuele Pizzuto ORCID: orcid.org/0000-0001-5776-654X⁵](#)

Nature volume **597**, pages 103–108 (2021)

- 14k Accesses
- 452 Altmetric
- [Metrics details](#)

Subjects

- [SARS-CoV-2](#)
- [Viral infection](#)
- [X-ray crystallography](#)

Abstract

The recent emergence of SARS-CoV-2 variants of concern^{1,2,3,4,5,6,7,8,9,10} and the recurrent spillovers of coronaviruses^{11,12} into the human population highlight the need for broadly neutralizing antibodies that are not affected by the ongoing antigenic drift and that can prevent or treat future zoonotic infections. Here we describe a human monoclonal antibody designated S2X259, which recognizes a highly conserved cryptic epitope of the receptor-binding domain and cross-reacts with spikes from all clades of

sarbecovirus. S2X259 broadly neutralizes spike-mediated cell entry of SARS-CoV-2, including variants of concern (B.1.1.7, B.1.351, P.1, and B.1.427/B.1.429), as well as a wide spectrum of human and potentially zoonotic sarbecoviruses through inhibition of angiotensin-converting enzyme 2 (ACE2) binding to the receptor-binding domain. Furthermore, deep-mutational scanning and in vitro escape selection experiments demonstrate that S2X259 possesses an escape profile that is limited to a single substitution, G504D. We show that prophylactic and therapeutic administration of S2X259 protects Syrian hamsters (*Mesocricetus auratus*) against challenge with the prototypic SARS-CoV-2 and the B.1.351 variant of concern, which suggests that this monoclonal antibody is a promising candidate for the prevention and treatment of emergent variants and zoonotic infections. Our data reveal a key antigenic site that is targeted by broadly neutralizing antibodies and will guide the design of vaccines that are effective against all sarbecoviruses.

[Download PDF](#)

Main

SARS-CoV-2 has caused a global pandemic with over 182 million infections and more than 3 million fatalities as of July 2021. SARS-CoV-2 genetic drift has resulted in emerging variants of concern that are characterized by higher transmissibility, immune evasion and disease severity relative to the ancestral isolate^{1,2,3,4,5,6,7,8,9,10}. Countermeasures, such as vaccines and therapeutic agents, are needed to cope with SARS-CoV-2 evolution and to protect against future sarbecovirus spillovers.

The sarbecovirus spike (S) glycoprotein mediates viral entry into host cells, represents the main target of neutralizing antibodies and is the focus of vaccine design¹¹. S comprises an S1 subunit, which recognizes host cell receptors, and an S2 subunit that promotes viral–cell membrane fusion^{12,13}. The S1 subunit includes the receptor-binding domain (RBD), which in the case of SARS-CoV and SARS-CoV-2 interacts with ACE2 to allow virus entry^{12,14,15,16}.

A highly conserved region on sarbecovirus RBDs, designated antigenic site II¹⁷, has previously been shown to elicit SARS-CoV and SARS-CoV-2 cross-neutralizing antibodies^{18,19}. However, site II becomes accessible only when at least two RBDs in the S trimer adopt an open conformation and thus has limited immunogenicity^{17,19}. Here we describe a site-II-targeting monoclonal antibody (mAb) designated S2X259 that possesses exceptional neutralization breadth within the *Sarbecovirus* subgenus (including all SARS-CoV-2 variants of concern), and an escape profile that is limited to a sole and rare substitution. In addition, we show that prophylactic and therapeutic administration of S2X259 protects Syrian hamsters against SARS-CoV-2 challenge with the prototypic virus and the B.1.351 variant of concern. Our findings identify

S2X259 as a promising countermeasure to protect against SARS-CoV-2 antigenic drift as well as potentially new zoonotic infections, and highlight the importance of RBD site II for the design of vaccines eliciting antibodies against all sarbecoviruses.

A broadly neutralizing sarbecovirus mAb

We identified the mAb designated S2X259 from the memory B cells of an individual who had recovered from COVID-19 (Supplementary Table 1); this mAb cross-reacted with 25 out of 26 S glycoproteins representative of all clades of sarbecovirus^{18,20} (Fig. 1a, Extended Data Fig. 1a–c). However, no binding was observed to embecovirus S (HCoV-HKU1 and HCoV-OC43) or merbecovirus S (MERS-CoV) glycoproteins (Extended Data Fig. 1a, b).

Fig. 1: Identification of a broadly neutralizing sarbecovirus mAb.

 **figure1**

a, Phylogenetic tree of sarbecovirus RBDs constructed using maximum likelihood analysis of amino acid sequences retrieved from GISAID and GenBank. Cross-reactivity within the *Sarbecovirus* subgenus for S2E12²¹, S309²² and ADG-2²⁰ is included for comparison. **b**, S2X259 binding to RBDs representative of the different sarbecovirus clades and SARS-CoV-2 variants as measured by ELISA. **c**, S2X259-mediated neutralization of SARS-CoV-2-Nluc authentic virus and SARS-CoV-2 S MLV-pseudotyped virus (MLV pp). **d**, **e**, S2X259-mediated neutralization of VSV pseudotypes containing SARS-CoV-2 S from isolates representing the B.1.1.7, B.1.351, P.1 and B.1.429 variants of concern (**d**) as well as single RBD mutants (**e**). **f**,

g, S2X259-mediated neutralization of VSV pseudotypes containing SARS-CoV-related (clade 1a) (**f**) or SARS-CoV-2-related (clade 1b) (**g**) S glycoproteins. $n = 2$ independent experiments. Error bars indicate s.d. of duplicates or triplicates.

The cross-reactivity of S2X259 within the *Sarbecovirus* subgenus was further confirmed using enzyme-linked immunosorbent assay (ELISA) with a panel of 12 RBDs (Fig. [1b](#)). S2X259 bound tightly to all of the RBDs tested (half maximal effective concentration (EC_{50}) of less than 32 ng ml $^{-1}$) with the exception of two bat strains, ZC45 (EC_{50} of 539 ng ml $^{-1}$) and Anlong-112 (EC_{50} of 1,281 ng ml $^{-1}$) (Fig. [1b](#)). The broad recognition of S from the divergent bat Southeast Asian and non-Southeast Asian clades (clades 2 and 3, respectively) highlights the exceptional breadth of S2X259 in comparison with previously reported mAbs [17,21,22](#) (Extended Data Fig. [1d](#)). Surface plasmon resonance indicated that the S2X259 Fab fragment bound to clade 1a, 1b and 3 RBDs with nano- to picomolar affinities, and to clade 2 RBDs with micro- to nanomolar affinities. Of note, the affinity of S2X259 for the RBD of the Guanxi pangolin coronavirus was 10–100-fold lower compared to other clade 1b RBDs (Extended Data Fig. [2a](#)). The S2X259 Fab recognized both the SARS-CoV-2 prefusion-stabilized S ectodomain trimer and the recombinant RBD with picomolar affinities, albeit with a slower on-rate for binding to S (presumably owing to limited epitope exposure) [12,13](#) (Extended Data Fig. [2a](#)). Finally, S2X259 binding was unaffected by several point mutants or constellation of mutations identified in circulating clinical isolates (K417V, N439K, Y453F and E484K) and in the RBDs of the B.1.1.7 (N501Y), B.1.351 (K417N/E484K/N501Y), B.1.427/B.1.429 (L452R) and P.1 (K417T/E484K/N501Y) lineages (Fig. [1b](#), Extended Data Fig. [2b](#)).

We next evaluated the neutralization potency of S2X259 using authentic SARS-CoV-2-Nluc as well as murine leukaemia virus (MLV) and vesicular stomatitis virus (VSV) pseudotyping systems. S2X259 potently neutralized authentic SARS-CoV-2, SARS-CoV-2 S MLV and SARS-CoV-2 S VSV with half-maximal inhibitory concentrations (IC_{50}) of 144 ng ml $^{-1}$, 87 ng ml $^{-1}$ and 213 ng ml $^{-1}$, respectively (Fig. [1c,d](#)). We subsequently evaluated S2X259-mediated neutralization against VSV pseudotypes containing SARS-CoV-2 S from the B.1.1.7, B1.351, P.1, B.1.427/B.1.429 and B.1.617.1 and other lineages, as well as the N439K and Y453F single mutations [5,23,24,25](#). S2X259 neutralized all of these mutants with a similar potency, consistent with the identical binding affinities determined using RBD mutants (Fig. [1d,e](#), Extended Data Figs. [2b](#), [3a,b](#), Supplementary Table [2](#)). Furthermore, S2X259 also neutralized a broad panel of ACE2-using sarbecoviruses with comparable potencies, including Guangdong pangolin coronavirus, RaTG13, SARS-CoV, WIV1, WIV16 and RsSHC014 (Fig. [1f,g](#), Supplementary Table [2](#)). Consistent with the surface plasmon resonance data, S2X259 neutralized VSV pseudotypes

containing S of Guanxi pangolin coronavirus but with reduced potency as compared to SARS-CoV-2 (Fig. 1g, Extended Data Fig. 2a, Supplementary Table 2).

Structural basis of sarbecovirus neutralization

Three-dimensional classification of the cryo-electron microscopy data showed the presence of SARS-CoV-2 S ectodomain trimers saturated with three S2X259 Fabs bound to open RBDs swung out to various extents, for which we determined a consensus structure at 3.2 Å resolution (Fig. 2a, b, Extended Data Fig. 4, Supplementary Table 3). We used local refinement to determine a 3.7 Å resolution map of the region corresponding to the S2X259 variable domains and RBD. In parallel, we crystallized the S2X259 Fab-bound SARS-CoV-2 RBD in the presence of the S2H97 Fab²⁶ and determined a 2.65 Å resolution X-ray structure (Fig. 2c–e, Supplementary Table 4).

Fig. 2: The S2X259 broadly neutralizing sarbecovirus mAb recognizes RBD antigenic site II.

 figure2

a, b, Cryo-electron microscopy structure of the prefusion SARS-CoV-2 S ectodomain trimer with three S2X259 Fab fragments bound to three open RBDs, viewed along two orthogonal orientations (Protein Data Bank (PDB) codes 7RA8 and 7RAL). **c,** The S2X259 binding mode involving contacts with multiple RBD regions. Residues corresponding to prevalent RBD mutations are shown as red spheres (PDB code 7M7W). **d, e,** Close-up views showing selected interactions formed between S2X259

and the SARS-CoV-2 RBD. In **a–e**, each SARS-CoV-2 S protomer is coloured distinctly (cyan, pink and gold), and the S2X259 light and heavy chain variable domains are coloured magenta and purple, respectively. N-linked glycans are rendered as blue spheres in **a–c**.

S2X259 recognizes a glycan-free, cryptic epitope within antigenic site IIa, which was previously defined on the basis of the nonclonally related S2X35 mAb that was isolated from the same donor¹⁷. S2X259 binding requires opening of two RBDs to grant access to the Fab in the context of the S trimer (Fig. [2a, b](#)). S2X259 contacts the RBD using both heavy and light chains, which respectively contribute two thirds and one third of the approximately 950 Å² paratope surface that is buried upon binding. Specifically, S2X259 uses complementary determining regions (CDRs) H1, H2, H3, L1 and L3 to contact RBD residues 369–386, which form two α-helices and an intervening β-strand belonging to the structurally conserved RBD β-sheet, as well as the residues 404–411 and 499–508, which form a continuous surface made up of an α-helix and a loop followed by an α-helix, respectively.

The S2X259 epitope is conserved in circulating SARS-CoV-2 isolates and does not comprise prevalent RBD mutations, such as S477N, N439K or L452R. S2X259 also circumvents residues 417 and 484, and contacts the backbone of residue N501 but not its side chain, explaining the high-affinity binding to the B.1.1.7, B.1.351 and P.1 RBDs and potent neutralization of pseudotyped variant of concern viruses (Fig. [1b, d](#) Extended Data Fig. [2b](#)). S2X259 broad neutralization results from the conservation of its epitope among sarbecoviruses and from the angle of approach of the Fab that circumvents the SARS-CoV N357 glycan present in all sarbecoviruses, except for SARS-CoV-2 (SARS-CoV-2 N370) (Extended Data Fig. [5](#)). The SARS-CoV-2 RBD and that of Guanxi pangolin coronavirus differ at only two positions (G504N and Y508F), and the structural data indicate that the former substitution would probably dampen mAb binding through steric hindrance as indicated by in vitro neutralization data (Fig. [1g](#)).

S2X259 binding to the RBD does not affect or prevent recognition by site-I or site-IV-targeting mAbs (Extended Data Fig. [6a, b](#)), which represent the majority of mAbs currently approved or in the clinic^{17,22,27,28}. Competition binding assays suggest that serum antibodies directed to the RBD antigenic site II are subdominant in individuals who have recovered from SARS-CoV-2 infection and vaccinated individuals, as compared to antibodies directed towards site I (Extended Data Fig. [6c](#)). Therefore, S2X259 can be used in combination with other site I or site IV mAbs to maximize breadth against currently circulating and emerging SARS-CoV-2 variants, as well as against future sarbecovirus infections from animal reservoirs.

S2X259 has an unusually narrow escape profile

Using deep-mutational scanning with a yeast-display RBD mutant library²⁹, we exhaustively mapped mutations that escape S2X259 binding. Notably, binding was strongly reduced by only a restricted number of amino acid substitutions, as compared to previously described neutralizing mAbs²⁶, with mutations at RBD site 504 yielding the most marked reduction in binding (Fig. 3a, Extended Data Fig. 7a–c). Indeed, substitution of G504 with most other amino acids reduced S2X259 binding compared to the wild-type (Wuhan-Hu-1) RBD, which highlights the importance of this residue for mAb recognition.

Fig. 3: S2X259 is resilient to a broad spectrum of escape mutations.

 figure3

a, Complete mapping of mutations reducing S2X259 binding using yeast-displayed RBD and deep-mutational scanning. Mean mutation effect on ACE2 affinity, RBD folding and contribution to S2X259 binding of substitutions at each position in the S2X259 epitope are reported. Asterisks denote mutations introducing N-linked glycosylation sites that may not be tolerated in full S. **b**, Frequency of mutants within the S2X259 epitope on the basis of SARS-CoV-2 genome sequences available on GISAID as of May 2021. S2X259 neutralizing activity against selected mutations is reported.

To further explore the functional importance of the deep-mutational scanning data, we passaged a replication-competent VSV-SARS-CoV-2 S chimera^{30,31} in Vero cells in the presence of S2X259 mAb. The selective pressure imposed by S2X259 led to the emergence of viral mutants containing the G504D mutation, which was the sole mutation present in each of the 18 neutralization-resistant plaques we sampled. This substitution did not impair viral fitness of the VSV-SARS-CoV-2 S chimera (Extended Data Fig. 7d,e, Supplementary Table 5). Under the same conditions, S2X35¹⁷ (which also targets antigenic site IIa) selected for seven escape mutants at five different positions (Y369S, K378E, D405N, V503D or V503G and G504D or G504S)²⁶ (Supplementary Table 5).

The G504D substitution is very rare: it has been detected in only 27 out of 1,345,708 (0.002%) SARS-CoV-2 genome sequences available on GISAID, as of May 2021 (Fig. 3b). Furthermore, in agreement with the deep-mutational scanning results, S2X259 efficiently neutralized VSV pseudoviruses containing S mutations located in S2X259-contact residues and found with high frequency in human isolates (Fig. 3b, Extended Data Fig. 7f). In the same analysis, we confirmed the loss of S2X259 neutralizing activity against the G504D mutant. S2X259 CDRL1 and CDRL3 contact G504 or surrounding residues, which explains why substituting a glycine with any other residue prevents mAb binding (via steric hindrance along with potential disruption of the interactions of the Fab with RBD residue D405, the latter being in close proximity to G504) (Extended Data Fig. 7g).

These data also support the hypothesis proposed above for the markedly reduced affinity and neutralization potency of S2X259 against Guanxi pangolin pseudovirus which contains the G504N mutation but maintains an aspartate at position 405 (Fig. 1g, Extended Data Figs. 2a, 7f, 8). However, S2X259 binding to clade 2 S glycoproteins that bear G504D or G504E mutations (for example, those of Shanxi2011 or YN2013) was not abrogated (Fig. 1b), possibly owing to substitution at the SARS-CoV-2 equivalent position 405 with a serine and to substantial deletions in the receptor-binding motif (Extended Data Fig. 8).

The finding that the only escape mutation for S2X259 is the rare G504D substitution suggests this mAb might have a high barrier for the emergence of SARS-CoV-2 escape mutants, which might prove essential during the next stages of the pandemic.

S2X259 protects hamsters from SARS-CoV-2 challenge

Consistent with structural data that indicates a clash between S2X259 and ACE2 for binding to the RBD (Extended Data Fig. 9a), we found that S2X259 efficiently blocked binding of SARS-CoV-2 and SARS-CoV RBDs to human ACE2 (Extended Data Fig. 9b).

In addition, S2X259 binding efficiently promoted shedding of the S1 subunit from cell-surface-expressed full-length SARS-CoV-2 S, as has previously been shown with other RBD-specific mAbs^{17,32,33,34,35} but not with S2M11 (which locks S in the prefusion closed state)²¹ (Extended Data Fig. 9c). The efficient S1 shedding induced by S2X259 might be responsible for the lack of Fc γ RIIa and Fc γ RIIIa activation, which we used as a proxy for antibody-dependent cellular phagocytosis and antibody-dependent cellular cytotoxicity, respectively (Extended Data Fig. 9d,e).

Indeed, performing the same assays using target cells that express an uncleavable prefusion stabilized SARS-CoV-2 S protein (unable to release the S1 subunit), we confirmed the ability of S2X259 mAb to induce activation of Fc γ RIIa and Fc γ RIIIa in these conditions (Extended Data Fig. 9f,g).

These data show that the primary mechanism of S2X259-mediated neutralization of SARS-CoV-2 and other sarbecoviruses relies on blocking viral attachment to host-cell receptors through competitive inhibition of ACE2 binding. Premature triggering of S1 subunit shedding could inactivate viruses before encountering target cells, albeit reducing activation of S2X259-mediated effector functions.

Next, we evaluated the prophylactic activity of S2X259 against challenge with the prototypic (Wuhan-Hu-1-related) SARS-CoV-2 in a Syrian hamster model³⁶. S2X259 was administered at 1, 5 and 25 mg per kg body weight via intraperitoneal injection before intranasal challenge, and the lungs of the hamsters were collected 4 days later for the quantification of viral RNA and replicating virus. Despite the lack of Fc-dependent effector functions *in vitro* and their proposed contribution to *in vivo* protection against SARS-CoV-2^{37,38}, administration of S2X259 at 1 mg kg⁻¹ reduced the amount of viral RNA and infectious virus detected in the lungs by approximately 2 and 2.5 orders of magnitude, respectively, compared to hamsters that received a control mAb (Fig. 4a,b). Furthermore, complete abrogation of virus replication was observed in the lungs of the hamsters that were treated with S2X259 at 5 and 25 mg kg⁻¹ relative to the control group (Fig. 4b). Therapeutic administration (after exposure) of S2X259 at 5 mg kg⁻¹ reduced the amounts of viral RNA detected by 3 log and abrogated viral replication in the lungs of all of the hamsters but one (Fig. 4c,d). Hamsters treated with S2X259 at 10 or 20 mg kg⁻¹ displayed a further approximately 1 log reduction of viral RNA and complete abrogation of viral replication (Fig. 4c,d). Viral replication in the lungs detected at day 4 inversely correlated with serum mAb concentration measured at the time of infection or at day 4 after infection in prophylactic and therapeutic settings, respectively (Extended Data Fig. 10a,b).

Fig. 4: S2X259 protects hamsters challenged with prototypic and B.1.351 SARS-CoV-2.

 **figure4**

a, b, Viral RNA loads (**a**) and replicating virus titres (50% tissue culture infectious dose (TCID₅₀)) (**b**) in the lungs of Syrian hamsters at 4 days after intranasal infection with prototypic SARS-CoV-2 (Wuhan-Hu-1-related) following prophylactic administration of S2X259 mAb at 1 ($n = 6$ hamsters), 5 ($n = 5$ hamsters) and 25 ($n = 5$

hamsters) mg kg⁻¹. **c, d**, Viral RNA loads (**c**) and replicating virus titres (TCID₅₀) (**d**) in the lungs of Syrian hamsters after therapeutic administration of S2X259 at 5 (*n* = 5 hamsters), 10 (*n* = 6 hamsters) and 20 (*n* = 6 hamsters) mg kg⁻¹ 24 h after infection with SARS-CoV-2 (Wuhan-Hu-1-related). **e, f**, Quantification of viral RNA loads (**e**) and replicating virus titres (TCID₅₀) (**f**) in the lungs of Syrian hamsters 4 days after intranasal challenge with B.1.351 SARS-CoV-2 variant of concern following prophylactic administration of S2X259 at 1 mg kg⁻¹ (*n* = 6 hamsters), 4 mg kg⁻¹ (*n* = 6 hamsters) and in combination with S309 (1 + 1 mg kg⁻¹, *n* = 6 hamsters). Data from one independent experiment are presented. Two-tailed Mann–Whitney test was used for statistical analysis of significance.

We next assessed the prophylactic activity of S2X259 against challenge with B.1.351 SARS-CoV-2³⁹. S2X259 at 1 and 4 mg kg⁻¹, or in combination with S309 (each mAb at 1 mg kg⁻¹), significantly decreased the amount of viral RNA detected in the lungs relative to hamsters that received a control mAb (Fig. 4e). Moreover, S2X259 at 4 mg kg⁻¹ or in combination with S309 (each mAb at 1 mg kg⁻¹) completely abrogated SARS-CoV-2 B.1.351 virus replication in most of the hamsters (Fig. 4f). As was the case for the prototypic SARS-CoV-2, the amount of infectious virus at day 4 detected in the group administered only with S2X259 also inversely correlated with serum mAb concentration measured at the time of infection (Extended Data Fig. 10c). Moreover, hamsters that received the mAb cocktail appeared to benefit from the additive contribution of the S309 mAb (Fig. 4f, Extended Data Fig. 10d, e).

Discussion

The large number of mutations in the S glycoprotein of SARS-CoV-2 variants of concern^{4,23,25} has been shown to abrogate or reduce the neutralization potency of several mAbs and vaccine-elicited antibodies^{5,6,7,8,9,10,40}. Moreover, the detection of a large pool of sarbecoviruses in bats and other mammals across several continents suggests that future cross-species transmission events are likely to occur^{41,42,43}. However, approved SARS-CoV-2 therapeutic agents and vaccines predominantly target or elicit immunity against immunodominant, but highly variable, epitopes and therefore have a limited efficacy against antigenic drift and genetically distinct strains with zoonotic potential^{40,44}.

Strategies for eliciting broadly neutralizing and protective sarbecovirus antibodies targeting the RBD have recently been proposed^{18,40}, and previous work has described mAbs that are now in the clinic with extended coverage within clade 1 sarbecoviruses^{20,22}. The recent results from a phase III study demonstrating efficacy (that is, 85% reduction of hospitalizations or deaths) of the cross-reactive derivative of the S309 mAb (sotrovimab) paves the way for the development of additional

neutralizing mAbs targeting conserved RBD epitopes that may broaden the coverage across sarbecoviruses and increase the barrier to resistance.

Here we identify the S2X259 mAb that targets the RBD antigenic site II and uniquely cross-reacts with all of the SARS-CoV-2 variants of concern that we evaluated, as well as with 30 S trimers or RBDs from all four sarbecovirus clades. S2X259 broadly neutralizes pseudoviruses containing S proteins from the B.1.1.7, B.1.351, P.1, B.1.427/B.1.429, and B.1.617.1 variants, and from other representative members of SARS-CoV-2 and SARS-CoV clades (clades 1b and 1a, respectively). Of note, S2X259 neutralizes pseudotyped WIV1 S and RsSHC014 S (clade 1a) viruses, which have previously been suggested to be potential zoonotic threats to human health^{41,42}. Moreover, S2X259 also recognizes S and RBDs from bat Southeast Asian and non-Southeast Asian sarbecovirus clades (clades 2 and 3, respectively). The coverage of clade 2 might be particularly important, as a previous study¹⁶ has shown that exogenous protease addition to pseudoviruses from this clade resulted in detectable S-mediated entry in human cells. Furthermore, although clade 2 and clade 3 sarbecoviruses have not been reported to bind human ACE2, mutation or recombination events could lead to human ACE2 binding and entry^{16,45}. Overall, these findings indicate that S2X259 is a promising candidate for clinical development and could be stockpiled as part of a pandemic preparedness toolbox.

On the basis of the growing body of data demonstrating that antibodies targeting the SARS-CoV-2 RBD account for most neutralizing activity in sera from patients who have recovered from COVID-19^{17,46}, we propose that RBD-based vaccines could better promote elicitation of high titres of S2X259-like neutralizing antibodies, owing to the enhanced accessibility of this antigenic site compared to S-based vaccines^{40,47}. Our serological competition data confirm that the antibody responses against the conserved site targeted by S2X259 mAb is subdominant in donors who have recovered from COVID-19 as well as in individuals who have received S-based mRNA vaccines. We anticipate these data will guide future efforts to develop vaccines that overcome the emergence of SARS-CoV-2 variants and that are effective against all sarbecoviruses.

Methods

Cell lines

Cell lines used in this study were obtained from ATCC (HEK293T, Vero and Vero E6), ThermoFisher Scientific (ExpiCHO-S cells, FreeStyle 293-F cells and Expi293F cells) or were generated via lentiviral transduction (ExpiCHO-S and HEK293T expressing

ACE2 (HEK293T-ACE2)). None of the cell lines used was authenticated. Cell lines were routinely tested for mycoplasma contamination.

Sample donors

Samples from an individual who had recovered from COVID-19 (designated as donor X (a 52-year-old man)) was obtained 75 days after symptom onset under study protocols approved by the local Institutional Review Boards (Canton Ticino Ethics Committee, Switzerland). Samples from additional nine individuals who had recovered from COVID-19 and nine vaccinated individuals were obtained under study protocols approved by the Canton Ticino Ethics Committee, Switzerland and the Ethical committee of Luigi Sacco Hospital, Milan. Samples from individuals who had recovered from COVID-19 were collected between 14 and 75 days after symptom onset; samples from vaccinated individuals were collected between 13 and 20 days after the second dose of an mRNA-based SARS-CoV-2 vaccine. All donors provided written informed consent for the use of blood and blood components (such as peripheral blood mononuclear cells (PBMCs), sera or plasma).

Isolation of PBMCs, plasma and sera

PBMCs were isolated from blood draw performed using tubes prefilled with heparin, followed by Ficoll density gradient centrifugation. PBMCs were either used freshly along SARS-CoV-2 S-protein-specific memory B cell sorting or stored in liquid nitrogen for later use. Sera were obtained from blood collected using tubes containing clot activator, followed by centrifugation and stored at -80 °C.

B cell isolation and recombinant S2X259 mAb production

Starting from freshly isolated PBMCs or upon cells thawing, B cells were enriched by staining with CD19 PE–Cy7 and incubation with anti-PE beads, followed by positive selection using LS columns. Enriched B cells were stained with anti-IgM, anti-IgD, anti-CD14 and anti-IgA, all PE labelled, and prefusion SARS-CoV-2 S with a biotinylated avi tag conjugated to streptavidin Alexa-Fluor 647 (Life Technologies). SARS-CoV-2 S-specific IgG⁺ memory B cells were sorted by flow cytometry via gating for PE-negative and Alexa-Fluor 647-positive cells. Sorted IgG⁺ memory B cells were seeded in 384-well microtitre plates on a monolayer of mesenchymal stromal cells at 0.5 cells per well in the presence CpG2006, IL-2, IL-6 and IL-21, as previously described⁴⁸. After 7 days, the supernatants were screened for binding to SARS-CoV-2 RBD and prefusion S as well as for neutralization activity using a high-throughput VSV SARS-CoV-2 S-based microneutralization assay. Selection of the antibody of interest in secondary screening was based on reactivity to a panel of sarbecovirus RBDs as well as on the neutralizing activity. Antibody VH and VL

sequences were obtained by RT-PCR and mAbs were expressed as recombinant human Fab fragment or as IgG1 (G1m3 allotype). ExpiCHO cells were transiently transfected with heavy and light chain expression vectors as previously described²².

Affinity purification was performed on ÄKTA Xpress FPLC (Cytiva) operated by UNICORN software version 5.11 (build 407) using HiTrap Protein A columns (Cytiva) for full length human and hamster mAbs and CaptureSelect CH1-XL MiniChrom columns (ThermoFisher Scientific) for Fab fragments, using PBS as mobile phase. Buffer exchange to the appropriate formulation buffer was performed with a HiTrap Fast desalting column (Cytiva). The final products were sterilized by filtration through 0.22-µm filters and stored at 4 °C.

Sarbecovirus sequence analysis and SARS-CoV-2 mutant counts

Alignment and phylogenetic tree of the strains within the *Sarbecovirus* subgenus was generated using MEGA X and CLC Main work bench 21.0.3 (Qiagen). The following sequence was retrieved from GISAID and NCBI: A021 (AAV97986.1). The viral sequences were obtained from GISAID EpiCoV project (<https://www.gisaid.org/>).

Analysis was performed on sequences submitted to GISAID up to 2 April 2021. The S protein sequences were either obtained directly from the protein dump provided by GISAID or, for the latest submitted sequences that were not yet incorporated in the protein dump on the day of data retrieval, from the genomic sequences with exonerate⁴⁹ 2.4.0–haf93ef1_3 (<https://quay.io/repository/biocontainers/exonerate?tab=tags>) using protein to DNA alignment with parameters -m protein2dna --refine full --minintron 999999 --percent 20 and using accession YP_009724390.1 as a reference. Multiple sequence alignment of all human S proteins was performed with mafft⁵⁰ 7.475–h516909a_0 (<https://quay.io/repository/biocontainers/mafft?tab=tags>) with parameters --auto --reorder --keeplength --addfragments using the same reference as above. S sequences that contained more than 10% ambiguous amino acids or that were less than 80% of the canonical protein length were discarded. A total of 923,686 sequences were used for analysis. Variants were then extracted as compared to the reference with R 4.0.2 (<https://www.r-project.org/>) using Biostrings 2.56.0.

Binding to cell-surface-expressed sarbecovirus S proteins by flow cytometry

ExpiCHO-S cells were seeded at 6×10^6 cells per ml in a volume of 5 ml in a 50-ml bioreactor. S-encoding plasmids HKU3 (QND76020.1), Rs3367 (AGZ48818.1), YN2013 (AIA62330.1), Rs4874 (ATO98205.1), Rs4255 (ATO98193.1), Rs4231 (ATO98157.1), Rs4084 (ATO98132.1), ZXC21 (AVP78042.1), ZC45 (AVP78031.1), Rp/Shaanxi2011 (AGC74165.1), Rm1/2004 (ABD75332.1), Rf1-2004 (ABD75323.1), BM48-31/BGR/2008 (YP_003858584.1), RaTG13 (QHR63300.2), PC4-127

(AAU93318.1), SARS-CoV-2 (YP_009724390.1), LYRa3 (AHX37569.1), Cp/Yunnan2011 (AGC74176.1), SARS-CoV Urbani (AAP13441.1), As6526 (ATO98108.1), BtKY72 (APO40579.1), RmYN02 (EPI_ISL_412977), Pangolin_Guangdong-2019 (EPI_ISL_410721), Pangolin-Guanxi-2017 (EPI_ISL_410539) and HKU3-12 (ADE34812.1) were diluted in cold OptiPRO SFM (Life Technologies, 12309-050), mixed with ExpiFectamine CHO Reagent (Life Technologies, A29130) and added to the cells. Transfected cells were then incubated at 37 °C with 8% CO₂ with an orbital shaking speed of 250 r.p.m. (orbital diameter of 25 mm) for 42 h. Transiently transfected ExpiCHO cells were collected and washed two times in wash buffer (PBS 2% FBS, 2 mM EDTA). Cells were counted and distributed into round-bottom 96-well plates (Corning, 3799) and incubated with S2X259 mAb serial dilutions from 50 µg ml⁻¹. Alexa Fluor647-labelled goat anti-human IgG secondary antibody (Jackson ImmunoResearch, 109-606-098) was prepared at 2 µg ml⁻¹ and added onto cells after two washing steps. Cells were then washed twice and resuspended in wash buffer for data acquisition at ZE5 cytometer (Biorad).

Recombinant protein production

Wild-type SARS-CoV-2 RBD (with N-terminal signal peptide and ‘ETGT’, and C-terminal 8×His tag) was expressed in Expi293F cells at 37 °C and 8% CO₂ in the presence of 10 µM kifunensine. Transfection was performed using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Cell culture supernatant was collected four days after transfection and supplemented with 10× PBS to a final concentration of 2.5× PBS (342.5 mM NaCl, 6.75 mM KCl and 29.75 mM phosphates). For crystallization, wild-type SARS-CoV-2 RBD was purified using a 5-ml HisTALON superflow cartridge (Takara Bio) followed by size-exclusion chromatography using a Superdex 200 Increase 10/300 GL column pre-equilibrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl. RBD was deglycosylated with EndoH and mixed with a 1.3-fold molar excess of S2X259 Fab and S2H97 Fab. The complex was purified on a Superdex 200 Increase 10/300 GL column pre-equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl. SARS-CoV-2 S hexapro protein, used for cryo-electron microscopy (cryo-EM) single particle studies, was expressed and purified as previously described²¹.

ELISA

Ninety-six half area well-plates (Corning, 3690) were coated overnight at 4 °C with 25 µl of sarbecovirus RBD proteins WIV1 (AGZ48831.1), Anlong-112 (ARI44804.1), YN2013 (AIA62330.1), SC2018 (QDF43815.1), ZC45 (AVP78031.1), Rp/Shaanxi2011 (AGC74165.1), BM48-31/BGR/2008 (YP_003858584.1), RaTG13 (QHR63300.2), SARS-CoV-2 (YP_009724390.1), SARS-CoV Urbani (AAP13441.1),

BtKY72 (APO40579.1), Pangolin_Guangdong-2019 (EPI_ISL_410721) and SARS-CoV-2 RBD mutants, prepared at $5 \mu\text{g ml}^{-1}$ in PBS pH 7.2. Plates were then blocked with PBS 1% BSA (Sigma-Aldrich, A3059) and subsequently incubated with mAb serial dilutions for 1 h at room temperature. After 4 washing steps with PBS 0.05% Tween 20 (PBS-T) (Sigma-Aldrich, 93773), goat anti-human IgG secondary antibody (Southern Biotech, 2040-04) was added and incubated for 1 h at room temperature. Plates were then washed four times with PBS-T and 4-nitrophenyl phosphate (pNPP, Sigma-Aldrich, 71768) substrate was added. After 30 min incubation, absorbance at 405 nm was measured by a plate reader (Biotek) and data were plotted using Prism GraphPad 9.1.0.

S2X259 blockade-of-binding to RBD

S2X259 and S2H14 mAbs were biotinylated using EZ-Link NHS-PEG solid phase biotinylation kit (ThermoFisher Scientific) and binding to RBD was tested to set optimal concentration to be used in the assay after sample desalting using Zeba Spin Desalting Columns (ThermoFisher Scientific). Half-area 96-well-plates were coated overnight at 4°C with SARS-CoV-2 RBD–mouse Fc Tag diluted at $1 \mu\text{g ml}^{-1}$ in PBS. After a blocking step with Blocker Casein (ThermoFisher Scientific), serial plasma dilutions in Blocker Casein were incubated 30 min at room temperature. Biotinylated S2X259 or S2H14 were added at a concentration achieving 80% of maximal binding and the mixture was incubated for 30 min at room temperature. Alkaline-phosphatase conjugated streptavidin (Jackson ImmunoResearch) was diluted at $0.5 \mu\text{g ml}^{-1}$ in Blocker Casein and added on plates previously washed 4 times with PBS 0.05% Tween 20. After 45 min incubation, plates were washed and 4-nitrophenyl phosphate substrate incubated for 1 h at room temperature. Absorbance at 405 nm was measured and percentage of inhibition was calculated as follows: $(1 - (\text{optical density (OD sample} - \text{OD negative control}) / (\text{OD positive control} - \text{OD negative control})) \times 100$.

MLV-based pseudotyped virus production and neutralization

To generate SARS-CoV-2 S MLV pseudotyped virus, HEK293T cells were seeded in 10-cm dishes in DMEM supplemented with 10% FBS. The next day cells were transfected with a SARS-CoV-2 S glycoprotein-encoding plasmid containing the D19 C-terminal truncation⁵¹, an MLV Gag-Pol packaging construct and the reporter vector pTG-Luc, using the X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's instructions. Cells were then incubated at 37°C with 5% CO₂ for 72 h. Supernatant was collected and cleared from cellular debris by centrifugation at 400g, and stored at -80°C .

For neutralization assays, Vero E6 cells were seeded into white 96-well plates (PerkinElmer) at 20,000 cells per well and cultured overnight at 37 °C with 5% CO₂ in 100 µl DMEM supplemented with 10% FBS and 1% penicillin–streptomycin. The next day, MLV-SARS-CoV-2 pseudovirus was activated with 10 µg ml⁻¹ TPCK-treated trypsin (Worthington Biochem) for 1 h at 37 °C. Recombinant antibodies at various concentrations were incubated with activated pseudovirus for 1 h at 37 °C. The Vero E6 cells were then washed with DMEM, and 50 µl of pseudovirus–mAbs mixes were added and incubated for 2 h at 37 °C with 5% CO₂. After incubation, 50 µl of DMEM containing 20% FBS and 2% penicillin–streptomycin were added and the cells were incubated 48 h at 37 °C with 5% CO₂. Following these 48 h of infection, culture medium was removed from the cells and 50 µl per well of Bio-Glo (Promega) diluted 1:2 with PBS with Ca²⁺Mg²⁺ (Thermo Fisher) was added to the cells and incubated in the dark for 15 min before reading on a Synergy H1 Hybrid Multi-Mode plate reader (Biotek). Measurements were done in duplicate, Relative light unit (RLU) values were converted to percentage of neutralization and plotted with a nonlinear regression curve fit in GraphPad Prism 9.1.0.

VSV-based pseudotype virus production and neutralization assay

SARS-CoV-2 S (CAD0240757.1), RaTG13 S (QHR63300.2), Pangolin-Guangdong S (QLR06867.1), Pangolin-Guanxi S (QIA48623.1), SARS-CoV S (YP 009825051.1), WIV1 S (AGZ48831.1), WIV16 S (ALK02457.1), RsSHCO14 S (AGZ48806.1), the variant of concern B.1.429 S (QTC60823) and the variants of concern, N437K, Y453F, B.1.1.7 S, B.1.351 S and P.1 S with their corresponding mutations inserted in the SARS-CoV-2 S (CAD0240757.1) backbone were used to pseudotype VSV. Pseudotyped viruses were prepared using 293T cells seeded in 10-cm dishes. In brief, cells in DMEM supplemented with 10% FBS, 1% penicillin–streptomycin were transfected with the plasmid encoding for the corresponding S glycoprotein using lipofectamine 2000 (Life Technologies) following manufacturer's indications. One day after transfection, cells were infected with VSV (G*ΔG-luciferase) and after 2 h, infected cells were washed 4 times with DMEM before adding medium supplemented with anti-VSV-G antibody (I1- mouse hybridoma supernatant diluted 1 to 50, from CRL- 2700, ATCC). Particles were collected 18 h after inoculation, clarified from cellular debris by centrifugation at 2,000g for 5 min and concentrated 10 times using a 30-kDa cut-off membrane and used for neutralization experiments, aliquoted and frozen at -80 °C until use in neutralization experiments.

For neutralization, stable 293T cells expressing ACE2⁵² in DMEM supplemented with 10% FBS, 1% penicillin–streptomycin were seeded at 40,000 cells per well into clear-bottom white-walled 96-well plates and cultured overnight at 37 °C. Twelve-point threefold serial dilutions of the corresponding mAb were prepared in DMEM and

pseudotyped VSV viruses were added 1:1 to each mAb dilution in the presence of anti-VSV-G antibody from I1- mouse hybridoma supernatant diluted 50 times. After 45 min incubation at 37 °C, 40 µl of the mixture was added to the cells and 2 h after infection, 40 µl DMEM was added to the cells. After 17–20 h 50 µl per well of One-Glo-EX substrate (Promega) was added to the cells and incubated in the dark for 5–10 min before reading on a Varioskan LUX plate reader (ThermoFisher). Measurements were done in duplicate with two independent productions of pseudotyped viruses and RLU values were converted to percentage of neutralization and plotted with a nonlinear regression curve fit in GraphPad Prism 9.1.0.

Generation of VSV SARS-CoV-2 S pseudotypes from a selection of circulating variants

Replication-defective VSV pseudovirus expressing SARS-CoV-2 S protein were generated as previously described⁵³ with some modifications. Plasmids encoding the SARS-CoV-2 S glycoprotein corresponding to the variants of concern used to produce SARS-CoV-2-VSV were obtained using a multispep overlap extension PCR protocol^{5,7}. In brief, the mutations of the different variant of concern lineages were encoded on each primer pair used to amplify sequential, overlapping fragments of the SARS-CoV-2 S encoding DNA sequence from a pCDNA_SARS-CoV-2_S D19 plasmid encoding the C-terminally truncated S protein that has been shown to be better expressed at the cell³⁰. The amplified overlapping fragments were separated on agarose gel and purified using Illustra GFX PCR DNA and Gel Band Purification Kit (Merck KGaA). The purified overlapping fragments were joined by performing 10 cycles of PCR without the addition of primers, followed by 30 PCR cycles using a single pair of external primers carrying 5' overhangs homologous to the vector backbone. The final PCR product was then cloned into the pCDNA3 vector using the Takara In-fusion HD cloning kit following manufacturer's instructions.

Lenti-X 293T cells (Takara, 632180) were seeded in 10-cm dishes at a density of 5×10^6 cells per cm² and the following day transfected with 10 µg of S expression plasmid with TransIT-Lenti (Mirus, 6600) according to the manufacturer's instructions. One day after transfection, cells were infected with VSV-luc (VSV-G) (Kerafast, EH1020-p.m.) for 1 h, rinsed three times with PBS, then incubated for an additional 24 h in complete medium at 37 °C. The cell supernatant was clarified by centrifugation, filtered (0.45 µm), aliquoted and frozen at –80 °C.

For VSV pseudovirus neutralization assays, Vero-E6 cells were grown in DMEM supplemented with 10% FBS and seeded into CulturePlate-96, White Opaque 96-well Microplate (PerkinElmer) at a density of 2×10^4 cells per well. The next day, mAbs were serially diluted in prewarmed complete medium, mixed at 1:1 ratio with pseudovirus and incubated for 1 h at 37 °C in round-bottom polypropylene plates.

Medium from cells was aspirated and 50 µl of virus–mAb complexes were added to cells and then incubated for 1 h at 37 °C. An additional 100 µl of prewarmed complete medium was then added on top of complexes and cells incubated for an additional 16–24 h. Conditions were tested in duplicate wells.

Virus–mAb-containing medium was then aspirated from cells and 50 µl of a 1:2 dilution of SteadyLite Plus (Perkin Elmer) in PBS with Ca²⁺Mg²⁺ was added to cells. Plates were incubated for 10 min at room temperature and then analysed on the Synergy H1 Hybrid Multi-Mode reader (Biotek). Measurements were done in two independent experiments and RLU values were converted to percentage of neutralization and plotted with a nonlinear regression curve fit with an upper constraint equal to 100 and a lower constraint equal to 0 in Prism (version 9.1.0). Half-maximal inhibitory concentration (IC₅₀) values across biological replicates are presented as arithmetic mean ± s.d.

Neutralization of SARS-CoV-2 B.1.351 pseudotyped VSV by S2X259 + S309 antibody cocktail

Vero E6 cells were plated into white 96-well plates (PerkinElmer) at 20,000 cells per well and cultured overnight at 37 °C 5% CO₂ in 100 µl culture medium (DMEM 10% FBS 1% penicillin–streptomycin). The next day, S2E12 and S309 antibodies were serially diluted in infection medium (DMEM 10% heat-inactivated FBS 1% penicillin–streptomycin) and mixed in a matrix format before incubation with SARS-CoV-2 B.1.351 pseudovirus (multiplicity of infection (MOI) 0.1) at 37 °C 5% CO₂ for 1 h. Cell culture medium was aspirated and 50 µl of mAbs–pseudovirus mix added onto cells. After incubation at 37 °C 5% CO₂ for 1 h, 100 µl of infection medium was added to cell and incubation continued for the following 20 h. For results read out, medium was removed and 50 µl per well of Steadylite PLUS (PerkinElmer) 1:2 diluted in PBS + Ca²⁺ and Mg²⁺ was added and incubated 15 min at room temperature in the dark. Luminescence signals were measured using a Synergy H1 reader (Biotek) with 1-s integration time. Data were analysed using MacSynergy II and synergy plots obtained with 99.9% confidence were used for graphical elaboration with GraphPad Prism 9.1.0.

Neutralization of authentic SARS-CoV-2-Nluc virus

Neutralization of authentic SARS-CoV-2 by entry-inhibition assay was determined using SARS-CoV-2-Nluc, an infectious clone of SARS-CoV-2 (based on strain 2019-nCoV/USA_WA1/2020) that encodes nanoluciferase in place of the viral ORF7 and demonstrated comparable growth kinetics to wild-type virus⁵⁴. Vero E6 cells were seeded into black-walled, clear-bottom 96-well plates at 2 × 10⁴ cells per well and

cultured overnight at 37 °C. The next day, 9-point fourfold serial dilutions of mAbs were prepared in infection medium (DMEM + 10% FBS). SARS-CoV-2-Nluc was diluted in infection medium at a final MOI of 0.01 plaque-forming units per cell, added to the mAb dilutions and incubated for 30 min at 37 °C. Medium was removed from the Vero E6 cells, mAb–virus complexes were added and incubated at 37 °C for 24 h. Medium was removed from the cells, Nano-Glo luciferase substrate (Promega) was added according to the manufacturer’s recommendations, incubated for 10 min at room temperature and the luciferase signal was quantified on a VICTOR Nivo plate reader (Perkin Elmer).

Affinity determination by surface plasmon resonance

Surface plasmon resonance binding measurements were performed using a Biacore T200 instrument, running buffer was Cytiva HBS-EP+ (pH 7.4), and all measurements were performed at 25 °C. For Fab binding, anti-AviTag pAb (for capturing S proteins) or StrepTactin XT (for capturing RBDs) were covalently immobilized on CM5 chips. S2X259 Fab concentrations were 11, 33, 100, and 300 nM run as single-cycle kinetics. Double reference-subtracted data were fit to a 1:1 binding model using Biacore Evaluation software, which yields an ‘apparent KD’ ($K_{D,\text{app}}$) for the S-binding data because the kinetics also reflect S conformational dynamics. For SARS-CoV-2 S the dissociation rate was too slow to fit, so the $K_{D,\text{app}}$ is reported as an upper limit. The K_D above 1 μM is approximate and was determined from a fit in which R_{\max} was set to a constant based on theoretical R_{\max} .

Affinity determination using biolayer interferometry

Biotinylated RBD (wild type, N501Y, K417N-E484K-N501Y, or K417T-E484K-N501Y) were immobilized at 5 ng μl^{-1} in undiluted 10X Kinetics Buffer (Pall) to SA sensors until a load level of 1.1 nm. A 1:3 dilution series of Fab in undiluted kinetics buffer starting at 10 nM was used for 600 s association before 600 s dissociation to determine protein–protein affinity. The data were baseline-subtracted and the plots fitted using the Pall FortéBio/Sartorius analysis software (version 12.0). Data were plotted in Prism 9.1.0.

Competition assay by biolayer interferometry

Biolayer interferometry was used to assess S2X259 competition with S309 and S2E12 using an Octet Red96 (ForteBio). All reagents were prepared in kinetics buffer (PBS 0.01% BSA) at the indicated concentrations. His-tagged SARS-CoV-2 RBD was prepared at 8 $\mu\text{g ml}^{-1}$ and loaded on prehydrated anti-penta-HIS biosensors (Sartorius) for 2.5 min. Biosensors were then moved into a solution containing S2X259 mAb and

association recorded for 7 min. A second association step was subsequently performed into S2X259 (as control), S309 and S2E12 mAbs solutions at 20 µg ml⁻¹ and recorded for 7 min. Response values were exported and plotted using GraphPad Prism 9.1.0.

Cell-surface mAb-mediated S1 shedding

CHO cells stably expressing wild-type SARS-CoV-2 S were resuspended in wash buffer (PBS 1% BSA, 2 mM EDTA) and treated with 10 µg ml⁻¹ TPCK-trypsin (Worthington Biochem) for 30 min at 37 °C. Cells were then washed and distributed into round bottom 96-well plates (90,000 cells per well). S2X259 was added to cells at 15 µg ml⁻¹ final concentration for 180 min at 37 °C. Cells were collected at different time points (5, 30, 60, 120 and 180 min), washed with wash buffer at 4 °C, and incubated with 1.5 µg ml⁻¹ secondary goat anti-human IgG, Fc fragment specific (Jackson ImmunoResearch) on ice for 20 min. Cells were washed and resuspended in wash buffer and analysed with ZE5 FACS (Biorad).

Measurement of Fc-effector functions

S2X259-dependent activation of human Fc γ RIIIa was performed with a bioluminescent reporter assay. ExpiCHO cells transiently expressing full-length wild-type SARS-CoV-2 S (target cells) or full-length prefusion stabilized SARS-CoV-2 S, which contains the 2P mutation and S1/S2 furin cleavage site mutation (RRARS to SGAG) as previously described¹², were incubated with different amounts of mAbs. After a 15-min incubation, Jurkat cells stably expressing Fc γ RIIIa receptor (V158 variant) or Fc γ RIIa receptor (H131 variant) and NFAT-driven luciferase gene (effector cells) were added at an effector to target ratio of 6:1 for Fc γ RIIIa and 5:1 for Fc γ RIIa. Signalling was quantified by the luciferase signal produced as a result of NFAT pathway activation. Luminescence was measured after 20 h of incubation at 37 °C with 5% CO₂ with a luminometer using the Bio-Glo Luciferase Assay Reagent according to the manufacturer's instructions (Promega).

In vivo mAb testing using a Syrian hamster model

KU LEUVEN R&D has developed and validated a SARS-CoV-2 Syrian hamster infection model^{36,39}. The SARS-CoV-2 (BetaCov/Belgium/GHB-03021/2020-EPI ISL 109 407976|2020-02-03), which is most closely related to the prototypic Wuhan-Hu-1 2019-nCoV (GenBank accession number MN908947.3) strain, and B.1.351 (hCoV105 19/Belgium/rega-1920/2021; EPI_ISL_896474, 2021-01-11) isolates used in this study were recovered from nasopharyngeal swabs taken from an asymptomatic returning traveller with RT-qPCR-confirmed SARS-CoV-2 infection in February 2020, and from a patient with respiratory symptoms in January 2021, respectively. A close

relatedness with the prototypic Wuhan-Hu-1 2019 SARS-CoV-2 and with B.1.351 lineage was confirmed by sequencing and phylogenetic analysis. Infectious viruses were isolated by serial passaging on Vero E6 cells and passage 6 for SARS-CoV-2 Wuhan-Hu-1 and passage 2 for B.1.351 viruses were used for the study. The titre of the virus stock was determined by end-point dilution on Vero E6 cells by a previously published method³⁵. This work was conducted in the high-containment A3 and BSL3+ facilities of the KU Leuven Rega Institute (3CAPS) under licences AMV 30112018 SBB 219 2018 0892 and AMV 23102017 SBB 219 20170589 according to institutional guidelines.

Syrian hamsters were purchased from Janvier Laboratories and were housed per two in ventilated isolator cages (IsoCage N Biocontainment System, Tecniplast) with ad libitum access to food and water and cage enrichment (wood block). Housing conditions and experimental procedures were approved by the ethical committee of animal experimentation of KU Leuven (licence P065-2020). Hamsters were randomized and the sample size was chosen on the basis of previous experience with this animal model. To evaluate the prophylactic efficacy 6–10-week-old female hamsters were administered by intraperitoneal injection with S2X259 mAb at 1, 4, 5 and 25 mg kg⁻¹ or a combination of S2X259 and S309 mAbs (1 mg kg⁻¹ + 1 mg kg⁻¹) 48 h before intranasal infection with 1.89×10^6 TCID₅₀ in 50 µl inoculum. Blood samples were collected before infection and serum was obtained for pharmacokinetics analysis. To assess the therapeutic efficacy, 6–8-week-old Syrian hamsters were administered by intraperitoneally with isotype control antibody (20 mg kg⁻¹) or S2X259 mAb (at 20, 10 or 5 mg kg⁻¹) 24 h after infection with SARS-CoV-2. Blood samples were collected when hamsters were killed, and serum was obtained for pharmacokinetics analysis. Hamsters were monitored for appearance, behaviour and weight. At day 4 after infection, hamsters were euthanized by intraperitoneal injection of 500 µl Dolethal (200 mg ml⁻¹ sodium pentobarbital, Vétoquinol SA). Lungs were collected, homogenized using bead disruption (Precellys) in 350 µl RLT buffer (RNeasy Mini kit, Qiagen) and centrifuged (10,000 r.p.m., 5 min, 4 °C) to pellet the cell debris. RNA was extracted using a NucleoSpin kit (Macherey-Nagel) according to the manufacturer's instructions. RT-qPCR was performed on a LightCycler96 platform (Roche) using the iTaq Universal Probes One-Step RTqPCR kit (BioRad) with N2 primers and probes targeting the nucleocapsid³⁶. Standards of SARS-CoV-2 cDNA (IDT) were used to express viral genome copies per mg tissue or per ml serum. To quantify infectious SARS-CoV-2 particles, end-point titrations were performed on confluent Vero E6 cells in 96-well plates. The hamster-derived samples for RNA quantification and viral load titration were run by technicians who were blinded to treatment group of the analysed samples. Viral titres were calculated by a previously published method³⁵ and were expressed as TCID₅₀ per mg tissue.

Evaluation of escape mutants using deep-mutational scanning

A previously described deep-mutational scanning approach⁵⁶ was used to identify RBD mutations that escape S2X259 binding. In brief, duplicate libraries containing virtually all possible amino acid changes compatible with ACE2 binding and RBD folding within the Wuhan-Hu-1 SARS-CoV-2 RBD sequence were expressed on the surface of yeast^{29,56}. Libraries were labelled at 59 ng ml⁻¹ S2X259 antibody (the 90% effective concentration for binding to yeast-displayed SARS-CoV-2 RBD determined in isogenic pilot binding experiments), and fluorescence-activated cell sorting (FACS) was used to select RBD⁺ cells that exhibit reduced antibody binding, as previously described^{56,57}. Libraries were sequenced before and after selection to determine permutation escape fractions, as previously described⁵⁷. Experiments were performed in duplicate with independently generated mutant libraries, and we report the average mutant escape fraction across the duplicates. Raw escape fractions are available on GitHub: [https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs](https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs/blob/main/results/supp_data/s2X259_raw_data.csv). Detailed steps of analysis and code for the deep mutational scanning selections are available on GitHub: https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs.

Selection of SARS-CoV-2 mAb escape mutants

VSV-SARS-CoV-2 chimera was used to select for SARS-CoV-2 S mAb-resistant mutants, as previously described^{30,31}. In brief, mAb-resistant mutants were recovered by plaque isolation on Vero cells with the indicated mAb in the overlay. The concentration of mAb in the overlay was determined by neutralization assays at a MOI of 100. Escape clones were plaque-purified on Vero cells in the presence of mAb, and plaques in agarose plugs were amplified on MA104 cells with the mAb present in the medium. Viral stocks were amplified on MA104 cells at an MOI of 0.01 in Medium 199 containing 2% FBS and 20 mM HEPES pH 7.7 (Millipore Sigma) at 34 °C. Viral supernatants were collected upon extensive cytopathic effect and clarified of cell debris by centrifugation at 1,000g for 5 min. Aliquots were maintained at -80 °C. Viral RNA was extracted from VSV-SARS-CoV-2 mutant viruses using RNeasy Mini kit (Qiagen), and S was amplified using OneStep RT-PCR Kit (Qiagen). The mutations were identified by Sanger sequencing (GENEWIZ). Their resistance was verified by subsequent virus infection in the presence or absence of antibody. In brief, Vero cells were seeded into 12-well plates for overnight. The virus was serially diluted using DMEM and cells were infected at 37 °C for 1 h. Cells were cultured with an agarose overlay in the presence or absence of mAb at 34 °C for 2 days. Plates were scanned on a biomolecular imager and expression of eGFP is show at 48 h after infection.

Crystallization, data collection, structure determination and analysis

Crystals of the SARS-CoV-2-RBD–S2X259–S2H97 Fab complex were obtained at 20 °C by sitting drop vapour diffusion. A total of 200 nl complex at 5.7 mg ml⁻¹ was mixed with 200 nl mother liquor solution containing 0.12 M monosaccharides mix, 20% (v/v) ethylene glycol, 10% (w/v) PEG 8000, 0.1 M Tris (base)/bicine pH 8.5, 0.02 M sodium chloride, 0.01 M MES pH 6 and 3% (v/v) Jeffamine ED-2003. Crystals were flash-frozen in liquid nitrogen. Data were collected at Beamline 9-2 of the Stanford Synchrotron Radiation Lightsource facility. Data were processed with the XDS software package⁵⁸ for a final dataset of 2.65 Å in space group P21. The RBD–S2X259–S2H97 Fab complex structure was solved by molecular replacement using phaser⁵⁹ from a starting model consisting of SARS-CoV-2 RBD (PDB code 7JX3) and homology models of the S2X259 and S2H97 Fabs built using the Molecular Operating Environment (MOE) software package from the Chemical Computing Group (<https://www.chemcomp.com>). Several subsequent rounds of model building and refinement were performed using Coot⁶⁰, ISOLDE⁶¹, Refmac5⁶² and MOE (<https://www.chemcomp.com>), to arrive at a final model for the ternary complex.

Cryo-EM sample preparation, data collection and data processing

Recombinantly expressed and purified Fab S2X259 and SARS-CoV-2 S hexapro were incubated at 1 mg ml⁻¹ with a 1.2 molar excess of Fab at 4 °C during 1 h. Three microlitres of the complex mixture were loaded onto freshly glow discharged R 2/2 UltrAuFoil grids (200 mesh) or lacey grids covered with a thin layer of manually added carbon, before plunge-freezing using a vitrobot MarkIV (ThermoFisher Scientific) with a blot force of 0 and 7–7.5 s blot time (for the UltrAuFoil grids) or with a blot force of −1 and 2.5 s blot time (for the lacey thin carbon grids) at 100% humidity and 21 °C.

Data were acquired on a FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. Automated data collection was carried out using Leginon⁶³ at a nominal magnification of 130,000 \times with a super-resolution pixel size of 0.525 Å. The dose rate was adjusted to 8 counts per pixel per s, and each movie was fractionated in 50 frames of 200 ms. Two datasets were collected from UltrAuFoil grids with the stage tilted 30° and 55° to circumvent particle preferential orientation. The third dataset was collected on lacey grids covered with a thin layer of carbon. The three datasets, with a total of 6,786 micrographs, were collected with a defocus range of between −0.8 and −2 μm. For each dataset, movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking and extraction were carried out using Warp⁶⁴. Particle images were extracted with a box size of 800 pixels² and binned to 400 yielding a pixel size of 1.05 Å. The three datasets were merged and two rounds of reference-free 2D classification were performed using cryoSPARC⁶⁵. Subsequently, one round of 3D classification with 50 iterations, using PDB code 6VXX as initial model, was carried out using Relion^{66,67} without imposing symmetry. Three-dimensional refinements were carried out using nonuniform refinement⁶⁸. Particle images from each dataset were subjected to Bayesian polishing⁶⁹ using Relion before merging them to perform another round of nonuniform refinement in cryoSPARC followed by per-particle defocus refinement and again nonuniform refinement. To improve the density of the S–S2X259 interface, the particles were symmetry-expanded and subjected to a Relion focus 3D classification without refining angles and shifts using a soft mask encompassing the RBD and S2X259 variable domains. Particles belonging to classes with the best-resolved local density were selected and subjected to local refinement using cryoSPARC. Local resolution estimation, filtering and sharpening were carried out using CryoSPARC. Reported resolutions are based on the gold-standard Fourier shell correlation of 0.143 criterion and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution⁷⁰.

Cryo-EM model building and analysis

UCSF Chimera⁷¹ and Coot⁶⁰ were used to fit atomic models (PDB code 6VXX or 6VYB) into the cryo-EM maps and the Fab variable domains were manually built. S2E12 was built in the locally refined map and subsequently validated using the Fab crystal structure. Models were refined and relaxed using Rosetta using both sharpened and unsharpened maps^{72,73}. Validation used Phenix⁷⁴, Molprobity⁷⁵ and Privateer⁷⁶. Figures were generated using UCSF ChimeraX⁷⁷.

Material availability

Materials generated in this study can be available on request and may require a material transfer agreement. The SARS-CoV-2 RBD mutant libraries (no. 1000000172) and unmutated parental plasmid (no. 166782) are available on Addgene

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All datasets generated and information presented in the study are available from the corresponding authors on reasonable request. The cryo-EM structure data and model are available from the PDB (and Electron Microscopy Data Bank) under accession codes [7RA8 \(EMD-24347\)](#) for the global refinement of the SARS-CoV-2 S and S2X259 complex and [7RAL \(EMD-24365\)](#) for the local refinement of the SARS-CoV-2 S RBD and S2X259 Fab variable domains. The X-ray structure data and model has been deposited at PDB code [7M7W](#) for RBD–S2H97–S2X259. Interactive escape maps and structural visualizations can be found at: https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Vir_mAbs/. Raw Illumina sequencing data from deep mutational scanning experiments are available on NCBI SRA, BioSample SAMN18315604 (SARS-CoV-2 mutant selection data). A complete table of deep-mutational scanning antibody escape fractions is provided on GitHub:

https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs/blob/main/results/supp_data/all_antibodies_raw_data.csv.

Code availability

A repository containing all code, analysis and summary notebooks for the analysis of the SARS-CoV-2 deep-mutational scanning escape selections is available on GitHub at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Vir_mAbs.

References

1. 1.
Deng, X. et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell* **184**, 3426–3437 (2021).
2. 2.
Challen, R. et al. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *Br. Med. J.* **372**, n579 (2021).
3. 3.
Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
4. 4.
Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).
5. 5.

McCallum, M. et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* <https://doi.org/10.1126/science.abi7994> (2021).

6. 6.

Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).

7. 7.

Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).

8. 8.

Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).

9. 9.

Chen, R. E. et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).

10. 10.

Madhi, S. A. et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B.1.351 variant. *N. Engl. J. Med.* **384**, 1885–1898 (2021).

11. 11.

Tortorici, M. A. & Veesler, D. Structural insights into coronavirus entry. *Adv. Virus Res.* **105**, 93–116 (2019).

12. 12.

Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292 (2020).

13. 13.

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

14. 14.

Kuhn, J. H., Li, W., Choe, H. & Farzan, M. Angiotensin-converting enzyme 2: a functional receptor for SARS coronavirus. *Cell. Mol. Life Sci.* **61**, 2738–2743 (2004).

15. 15.

Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).

16. 16.

Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).

17. 17.

Piccoli, L. et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042 (2020).

18. 18.

Cohen, A. A. et al. Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021).

19. 19.

Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).

20. 20.

Rappazzo, C. G. et al. Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. *Science* **371**, 823–829 (2021).

21. 21.

Tortorici, M. A. et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* **370**, 950–957 (2020).

22. 22.

Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).

23. 23.

Davies, N. G. et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).

24. 24.

Thomson, E. C. et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187 (2021).

25. 25.

Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).

26. 26.

Starr, T. N. et al. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* <https://doi.org/10.1038/s41586-021-03807-6> (2021).

27. 27.

Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

28. 28.

Cathcart, A L. et al. The dual function monoclonal antibodies VIR-7831 and VIR-7832 demonstrate potent in vitro and in vivo activity against SARS-CoV-2. Preprint at <https://doi.org/10.1101/2021.03.09.434607> (2021).

29. 29.

Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 (2020).

30. 30.

Case, J. B. et al. Replication-competent vesicular stomatitis virus vaccine vector protects against SARS-CoV-2-mediated pathogenesis in mice. *Cell Host Microbe* **28**, 465–474 (2020).

31. 31.

Liu, Z. et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **29**, 477–488 (2021).

32. 32.

Huo, J. et al. Neutralization of SARS-CoV-2 by destruction of the prefusion spike. *Cell Host Microbe* **28**, 445–454.e6 (2020).

33. 33.

Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039 (2019).

34. 34.

Wee, A. Z. et al. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* **369**, 731–736 (2020).

35. 35.

Wrobel, A. G. et al. Antibody-mediated disruption of the SARS-CoV-2 spike glycoprotein. *Nat. Commun.* **11**, 5337 (2020).

36. 36.

Boudewijns, R. et al. STAT2 signaling restricts viral dissemination but drives severe pneumonia in SARS-CoV-2 infected hamsters. *Nat. Commun.* **11**, 5838 (2020).

37. 37.

Schäfer, A. et al. Antibody potency, effector function, and combinations in protection and therapy for SARS-CoV-2 infection in vivo. *J. Exp. Med.* **218**, e20201993 (2021).

38. 38.

Winkler, E. S. et al. Human neutralizing antibodies against SARS-CoV-2 require intact Fc effector functions and monocytes for optimal therapeutic protection. *Cell* **184**, 1804–1820 (2021).

39. 39.

Abdelnabi, R. et al. Comparing infectivity and virulence of emerging SARS-CoV-2 variants in Syrian hamsters. *EBioMedicine* **68**, 103403 (2021).

40. 40.

Walls, A. C. et al. Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines. Preprint at <https://doi.org/10.1101/2021.03.15.435528> (2021).

41. 41.

Menachery, V. D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).

42. 42.

Menachery, V. D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl Acad. Sci. USA* **113**, 3048–3053 (2016).

43. 43.

Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).

44. 44.

Martinez, D. R. et al. Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice.
Science <https://doi.org/10.1126/science.abi4506> (2021).

45. 45.

Makarenkov, V., Mazoure, B., Rabusseau, G. & Legendre, P. Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *BMC Ecol. Evol.* **21**, 5 (2021).

46. 46.

Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by

polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 (2021).

47. 47.

Walls, A. C. et al. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382 (2020).

48. 48.

Pinto, D. et al. Structural basis for broad HIV-1 neutralization by the MPER-specific human broadly neutralizing antibody LN01. *Cell Host Microbe* **26**, 623–637 (2019).

49. 49.

Slater, G. S. C & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).

50. 50.

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

51. 51.

Ou, X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, 1620 (2020).

52. 52.

Crawford, K. H. D. et al. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* **12**, 513 (2020).

53. 53.

Takada, A. et al. A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94**, 14764–14769 (1997).

54. 54.

Xie, X. et al. A nanoluciferase SARS-CoV-2 for rapid neutralization testing and screening of anti-infective drugs for COVID-19. *Nat. Commun.* **11**, 5214 (2020).

55. 55.

Reed, L. J. & Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Epidemiol.* **27**, 493–497 (1938).

56. 56.

Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).

57. 57.

Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).

58. 58.

Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).

59. 59.

McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

60. 60.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

61. 61.

Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).

62. 62.

Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).

63. 63.

Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

64. 64.

Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).

65. 65.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

66. 66.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

67. 67.

Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).

68. 68.

Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

69. 69.

Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

70. 70.

Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).

71. 71.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

72. 72.

Frenz, B. et al. Automatically fixing errors in glycoprotein structures with Rosetta. *Structure* **27**, 134–139 (2019).

73. 73.

Wang, R. Y. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).

74. 74.

Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta*

Crystallogr. D **75**, 861–877 (2019).

75. 75.

Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

76. 76.

Agirre, J. et al. Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).

77. 77.

Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

78. 78.

Pinto, D. et al. A human antibody that broadly neutralizes betacoronaviruses protects against SARS-CoV-2 by blocking the fusion machinery. Preprint at <https://doi.org/10.1101/2021.05.09.442808> (2021).

Acknowledgements

We thank H. Tani for providing the reagents necessary for preparing VSV pseudotyped viruses; C. Castado and N. Blais for their help in the selection of the genetically divergent sarbecoviruses used in this study; J. C. Nix for X-ray data collection; I. Hoffman and T. I. Croll for assistance in refinement of crystal structures; and J. Dillen for assistance with surface plasmon resonance data collection. This study was supported by the National Institute of General Medical Sciences (R01GM120553 to D.V.), the National Institute of Allergy and Infectious Diseases (DP1AI158186 and HHSN272201700059C to D.V.; R01AI141707 to J.D.B.), a Pew Biomedical Scholars Award (D.V.), Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (D.V.), Fast

Grants (D.V.), the Pasteur Institute (M.A.T), the Damon Runyon Cancer Research Foundation (T.N.S.) and the University of Washington Arnold and Mabel Beckman cryo-EM centre. J.D.B. is an Investigator of the Howard Hughes Medical Institute. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Author information

Author notes

1. These authors contributed equally: M. Alejandra Tortorici, Nadine Czudnochowski, Tyler N. Starr, Roberta Marzi

Affiliations

1. Department of Biochemistry, University of Washington, Seattle, WA, USA

M. Alejandra Tortorici, Alexandra C. Walls, John E. Bowen, Zhaoqian Wang, Samantha K. Zepeda & David Veesler

2. Institut Pasteur and CNRS UMR 3569, Unité de Virologie Structurale, Paris, France

M. Alejandra Tortorici

3. Vir Biotechnology, San Francisco, CA, USA

Nadine Czudnochowski, Julia Di Julio, Michael P. Housley, Florian A. Lempp, Laura E. Rosen, Exequiel Dellota Jr, Hannah Kaiser, Martin

Montiel-Ruiz, Jiayi Zhou, Barbara Guarino, Colin Havenar-Daughton, Amalio Telenti, Herbert W. Virgin & Gyorgy Snell

4. Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Tyler N. Starr, Amin Addetia & Jesse D. Bloom

5. Humabs Biomed SA, a subsidiary of Vir Biotechnology, Bellinzona, Switzerland

Roberta Marzi, Fabrizia Zatta, Stefano Jaconi, Anna De Marco, Dora Pinto, Martina Beltramello, Istvan Bartha, Katja Culap, Nicole Sprugasci, Christian Saliba, Eneida Vetti, Isabella Giacchettosasselli, Chiara Silacci Fregni, Michael A. Schmid, Fabio Benigni, Elisabetta Cameroni, Davide Corti & Matteo Samuele Pizzuto

6. Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO, USA

Zhuoming Liu & Sean P. J. Whelan

7. Rega Institute for Medical Research, Laboratory of Virology and Chemotherapy, KU Leuven, Leuven, Belgium

Rana Abdelnabi, Shi-Yan Caroline Foo & Johan Neyts

8. Howard Hughes Medical Institute, Seattle, WA, USA

Jesse D. Bloom

Contributions

Experiment design was performed by M.A.T., N.C., T.N.S., R.M., A.C.W., S.P.J.W., J.D.B., D.C., D.V. and M.S.P.; E.C. and F.B. recruited donors and collected samples; R.M., F.Z., A.D.M., D.P. and M.B. processed PBMC samples; M.A.T., N.C., S.J., A.C.W., J.E.B., K.C., N.S., C.S., I.G.-S. and

E.C. performed protein expression and purification; R.M., F.Z., A.D.M., D.P., M.B. and C.S.F. isolated mAbs; M.A.T., R.M., A.C.W., F.Z., D.P., M.B., M.P.H., F.A.L., E.D. Jr, H.K., M.M.-R., J.Z., S.K.Z. and C.H.-D. performed binding and neutralization assays; A.C.W., F.Z., A.D.M. and L.E.R. undertook biolayer interferometry and surface plasmon resonance assays; M.A.T., Z.W. and D.V. collected and processed cryo-EM data, and built the models; N.C. and G.S. performed crystallization experiments and determined the X-ray structure; D.P. and A.D.M. performed the ACE2 binding inhibition and S1 shedding; B.G. evaluated effector functions; T.N.S., A.A. and J.D.B. performed deep-mutational scanning; J.D.I., I.B. and A.T. undertook bioinformatic analysis of virus diversity and variants; Z.L. and S.P.J.W. selected and sequenced escape mutants; R.A., S.-Y.C.F., E.V., M.A.S., F.B., J.N., D.C. and M.S.P. performed hamster model experiments and data analysis; M.A.T., N.C., T.N.S., R.M., A.C.W., J.D.B., D.C., D.V. and M.S.P. analysed data; M.A.T., N.C., T.N.S., R.M., J.N., A.T., G.S., H.W.V., S.P.J.W., J.D.B., D.C., D.V. and M.S.P. wrote the manuscript; F.A.L., L.E.R., C.H.-D., M.A.S., F.B., E.C., J.N., A.T., G.S., H.W.V., S.P.J.W., J.D.B., D.C., D.V. and M.S.P. supervised the project; and J.D.B. and D.V. acquired funding.

Corresponding authors

Correspondence to [Davide Corti](#) or [David Veesler](#) or [Matteo Samuele Pizzuto](#).

Ethics declarations

Competing interests

N.C., R.M., F.Z., S.J., J.D.I., A.D.M., D.P., M.B., I.B., M.H., F.A.L., L.E.R., E.D. Jr, H. K., M.M.-R., J.Z., K.C., N.S., C.S., E.V., I.G.-S., C.S.F., C.H.-D., M.A.S., F.B., E.C., A.T., G.S., H.W.V., D.C. and M.S.P. are employees of Vir Biotechnology Inc. and may hold shares in Vir Biotechnology Inc. D.C. is currently listed as an inventor on multiple patent applications, which disclose the subject matter described in this Article. The laboratory of J.N. has received sponsored research agreements from Vir Biotechnology Inc.

After the submission of the initial version of this study, J.D.B. began consulting for Moderna on viral evolution and epidemiology. J.D.B. has the potential to receive a share of IP revenue as an inventor on a Fred Hutchinson Cancer Research Center-optioned technology and patent (application WO2020006494) related to deep-mutational scanning of viral proteins. H.W.V. is a founder of PierianDx and Casma Therapeutics. Neither company provided funding for this work or is performing related work. D.V. is a consultant for Vir Biotechnology Inc. The laboratory of D.V. has received a sponsored research agreement from Vir Biotechnology Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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 Breadth of S2X259 mAb across the Sarbecovirus subgenus.](#)

a, Flow cytometry analysis of S2X259 cross-reactivity with a panel of 26 S glycoproteins representative of sarbecovirus clades 1a, 1b, 2 and 3. The colours represent the lowest concentration of mAb at which binding was observed. **b**, FACS binding of S2X259 to a panel of sarbecovirus S glycoproteins transiently expressed in ExpiCHO cells. Results represent the binding frequency normalized using the ratio between mean fluorescence intensities for S2X259 and an expression control mAb⁷⁸. **c**, FACS gating strategy used to assess binding of S2X259 or of the expression control mAb. **d**, ELISA binding of site-I-targeting S2E12²¹, site-II-targeting S2X259 and S2X35¹⁷, and site-IV-targeting S309²² mAbs to RBDs

spanning the different clades of the *Sarbecovirus* subgenus. EC₅₀s are reported. $n = 2$ independent experiments.

Extended Data Fig. 2 S2X259 Fab binding to recombinant sarbecovirus RBDs, prefusion SARS-CoV-2 S ectodomain trimer and RBD variants.

a, S or RBD antigens were captured on the sensor chip surface and binding to S2X259 Fab at 11, 33, 100 and 300 nM was monitored successively, in single-cycle kinetics format, by surface plasmon resonance. All data have been fit to a 1:1 binding model (black dashed line) and the equilibrium dissociation constant (K_D) is reported. For the S-binding data, we report an apparent K_D ($K_{D,app}$) as kinetics are affected by conformational dynamics between open and closed RBD states. The colouring scheme matches the phylogenetic tree in Fig. 1a. **b**, Biolayer interferometry binding analysis of the S2X259 Fab to wild-type or variant-of-concern SARS-CoV-2 biotinylated RBDs immobilized at the surface of SA biosensors. The data are coloured according to the key and fits to a 1:1 binding model are shown as black lines. Equilibrium dissociation constants (K_D) are reported above each plot.

Extended Data Fig. 3 S2X259 neutralizing activity is not affected by mutations present in circulating SARS-CoV-2 variant of concerns and recent lineages.

a, b, Neutralization by S2X259 of SARS-CoV-2 S VSV pseudotyped circulating variants shown as IC₅₀ (a) and IC₅₀ fold change as compared to Wuhan-Hu-1 virus with D614G substitution (b). Schematic of the SARS-CoV-2 S and the mutation landscape in each variant is shown beneath the panels. Del, deletion. Asterisks denotes virus that was tested as single RBD mutant in Fig. 1e.

Extended Data Fig. 4 Cryo-EM data processing and validation of S2X259-bound SARS-CoV-2 S.

a, b, Representative electron micrographs (**a**) and class averages (**b**) of SARS-CoV-2 S in complex with the S2X259 Fab. Scale bars, 500 Å (**a**), 100 Å (**b**). **c**, Gold-standard Fourier shell correlation curves for the S trimer bound to three S2X259 Fabs (solid black line) and the locally refined reconstruction of the RBD–S2X259 variable domains (dashed black line). The 0.143 cut-off is indicated by a horizontal dashed grey line. **d, e**, Local resolution map for the open S trimer bound to three S2X259 Fabs (**d**) and the locally refined reconstruction of the RBD–S2X259 variable domains (**e**). **f**, Cryo-EM data processing flowchart.

Extended Data Fig. 5 The S2X259 angle of approach for binding to the SARS-CoV-2 RBD allows S2X259 to circumvent the SARS-CoV N357 glycan present in all sarbecovirus RBDs except SARS-CoV-2.

Ribbon diagram showing a superimposition of the S2X259-bound and S2A4-bound (PDB code 7JVA) SARS-CoV-2 RBD¹⁷. The SARS-CoV glycan at position N357 was modelled on the basis of the S230-bound SARS-CoV S structure (PDB code 6NB6³³) and is predicted to sterically hinder S2A4 binding (red star) but not S2X259 binding. The mAb light and heavy chains are coloured magenta and purple (S2X259) or light and dark green (S2A4), respectively. N-linked glycans are rendered as blue spheres.

Extended Data Fig. 6 S2X259 can bind to the SARS-CoV-2 RBD in presence of site-I and -IV-targeting mAbs.

a, View of site-I-targeting S2E12²¹ (pink), site-II-targeting S2X259 (magenta) and site-IV-targeting S309²² (purple) mAb bound to the SARS-CoV-2 RBD (light blue). **b**, Competition binding assays for S2X259 versus site-I-targeting S2E12 and site-IV-targeting S309 mAbs on SARS-CoV-2 RBD as measured by biolayer interferometry. One independent experiment out of two is shown. **c**, Competition ELISA (blockade-of-binding) between site-I-targeting S2H14¹⁷ or site-II-targeting S2X259 and sera or plasma from individuals who had recovered from COVID-19 ($n = 10$ biological samples) and vaccinated individuals ($n = 9$ biological samples). Each plot shows the magnitude of inhibition of binding to immobilized RBD in the

presence of each mAb, expressed as reciprocal sera or plasma dilution blocking 80% of the maximum binding response. The individual from whom S2X259 mAb was isolated is shown as a square. Statistical analysis was performed using two-tailed Mann–Whitney test.

Extended Data Fig. 7 S2X259 has a high barrier for the emergence of resistance mutants.

a, FACS gates used in deep-mutational scanning to select escape variants. Top, yeast controls expressing unmutated SARS-CoV-2 RBD labelled at relevant S2X259 concentrations for setting of selection gates. Bottom, fraction of cells in SARS-CoV-2 mutant libraries falling into the antibody-escape bin. **b**, Correlation in site-level (top, sum of escape fractions for mutations at a site) and mutation-level (bottom) escape between independently generated and assayed RBD mutant libraries. **c**, Line plot of escape mutants along all positions in the SARS-CoV-2 RBD (left). Pink lines indicate sites that escape S2X259 binding illustrated at the mutation-level in logoplots (right). In logoplots, the height of a letter scales with its escape fraction. Letters are coloured according to their deleterious consequences for ACE2 binding (middle) or RBD expression (right) as determined in previous deep-mutational scanning experiments²⁹. **d**, Plaque assay using VSV-SARS-CoV-2 chimeric virus on Vero cells with no mAb (−) or S2X259 (+) in the overlay to isolate escape mutants (red arrow). Data are representative of three independent experiments. **e**, Plaque assays performed to validate the VSV-SARS-CoV-2 G504D mutant in Vero cells in the presence (+) or absence (−) of S2X259 in the overlay. Representative image of two independent experiments is shown. **f**, S2X259 in vitro neutralizing activity against SARS-CoV-2 S VSV pseudotyped mutants. For each mutant the fold change of the IC₅₀ geometric mean versus SARS-CoV-2 S D614G is reported. *Q506K displayed a tenfold reduction in viral entry in comparison to the other mutants. Results from two independent experiments are reported. **g**, Zoomed-in view of the S2X259–RBD interface showing that the G504D substitution would disrupt mAb binding owing to steric hindrance (indicated with a red star).

Extended Data Fig. 8 S2X259 epitope conservation across sarbecovirus clades.

Protein sequence alignment of representative sarbecovirus RBDs with strictly conserved residues shown as dots. Overall conservation is represented as a bar plot at the bottom of the figure. Residue positions are based on SARS-CoV-2. Residues determined to be most important for S2X259 binding are denoted in dark green. Substitutions at positions D405 and G504 are indicated in pink and blue or orange, respectively. Additional residues representing the extended epitope are denoted in grey. Different clades within the *Sarbecovirus* subgenus are overlaid in grey (clade 1a), red (clade 1b), green (clade 2) and light blue (clade 3).

Extended Data Fig. 9 Inhibition of ACE2 engagement, S1 subunit shedding and activation of Fc γ RIIa and Fc γ RIIIa in vitro.

a, S2X259 (purple and pink) and ACE2 (dark green) bind partially overlapping binding sites on the SARS-CoV-2 RBD (blue). **b**, Pre-incubation of serial dilutions of S2X259 with SARS-CoV-2 (red) or the SARS-CoV (black) RBDs prevents binding to immobilized human ACE2 (hACE2) ectodomain in ELISA. **c**, mAb-mediated S1 subunit shedding from cell-surface-expressed SARS-CoV-2 S as determined by flow cytometry. S2E12 was included as positive control and S2M11 was included as negative control. **d**, **e**, NFAT-driven luciferase signal induced in Jurkat cells stably expressing Fc γ RIIa H131 (**d**) variant or Fc γ RIIIa V158 (**e**) variant by S2X259 binding to full-length wild-type SARS-CoV-2 S on ExpiCHO target cells. **f**, **g**, NFAT-driven luciferase signal induced in Jurkat cells stably expressing Fc γ RIIa H131 (**f**) or Fc γ RIIIa V158 (**g**) variants by S2X259 binding to uncleavable full-length prefusion stabilized SARS-CoV-2 S (unable to release the S1 subunit) transiently expressed in ExpiCHO cells. SE12, S2M11, S309 and S309-GRLR mAbs are included as controls.

Extended Data Fig. 10 Correlation between mAb concentration and infectious virus in vivo and in vitro neutralizing activity of S2X259 + S309 antibody cocktail.

a–c, Infectious virus titres in the lungs at 4 days after infection plotted as a function of serum mAb concentrations before infection (day 0) with prototypic SARS-CoV-2 (**a**) and SARS-CoV-2 B.1.351 (**c**) in prophylactic setting or at 4 days after infection with prototypic SARS-CoV-2 (**b**) in therapeutic setting. **d, e**, SARS-CoV-2 B1.351 VSV-based pseudotype neutralization (**d**) and synergy score (**e**) measured combining S2X259 and S309 mAbs. Synergy was calculated using MacSynergyII (synergy score: 0.98, antagonism score: 0). Results represent four replicates of one independent experiment.

Supplementary information

Supplementary Tables

This file contains Supplementary Tables 1-5.

Reporting Summary

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Tortorici, M.A., Czudnochowski, N., Starr, T.N. *et al.* Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* **597**, 103–108 (2021). <https://doi.org/10.1038/s41586-021-03817-4>

- Received: 29 March 2021
- Accepted: 09 July 2021
- Published: 19 July 2021

- Issue Date: 02 September 2021
 - DOI: <https://doi.org/10.1038/s41586-021-03817-4>
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03817-4>

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

- Article
- Open Access
- [Published: 14 July 2021](#)

cGAS-like receptors sense RNA and control 3' 2'-cGAMP signalling in *Drosophila*

- [Kailey M. Slavik](#)^{1,2},
- [Benjamin R. Morehouse](#) ORCID: orcid.org/0000-0003-3352-5463^{1,2},
- [Adelyn E. Ragucci](#)^{1,2},
- [Wen Zhou](#)^{1,2 nAff9},
- [Xianlong Ai](#)³,
- [Yuqiang Chen](#)³,
- [Lihua Li](#)³,
- [Ziming Wei](#)³,
- [Heike Bähre](#)⁴,
- [Martin König](#)⁴,
- [Roland Seifert](#)^{4,5},
- [Amy S. Y. Lee](#)^{2,6},
- [Hua Cai](#)³,
- [Jean-Luc Imler](#)^{3,7} &
- [Philip J. Kranzusch](#) ORCID: orcid.org/0000-0002-4943-733X^{1,2,8}

Nature volume 597, pages 109–113 (2021)

- 10k Accesses
- 85 Altmetric
- [Metrics details](#)

Subjects

- [Innate immunity](#)
- [Structural biology](#)

Abstract

Cyclic GMP–AMP synthase (cGAS) is a cytosolic DNA sensor that produces the second messenger cG[2'–5']pA[3'–5']p (2'3'-cGAMP) and controls activation of innate immunity in mammalian cells^{1,2,3,4,5}. Animal genomes typically encode multiple proteins with predicted homology to cGAS^{6,7,8,9,10}, but the function of these uncharacterized enzymes is unknown. Here we show that cGAS-like receptors (cGLRs) are innate immune sensors that are capable of recognizing divergent molecular patterns and catalysing synthesis of distinct nucleotide second messenger signals. Crystal structures of human and insect cGLRs reveal a nucleotidyltransferase signalling core shared with cGAS and a diversified primary ligand-binding surface modified with notable insertions and deletions. We demonstrate that surface remodelling of cGLRs enables altered ligand specificity and used a forward biochemical screen to identify cGLR1 as a double-stranded RNA sensor in the model organism *Drosophila melanogaster*. We show that RNA recognition activates *Drosophila* cGLR1 to synthesize the novel product cG[3'–5']pA[2'–5']p (3'2'-cGAMP). A crystal structure of *Drosophila* stimulator of interferon genes (dSTING) in complex with 3'2'-cGAMP explains selective isomer recognition, and 3'2'-cGAMP induces an enhanced antiviral state in vivo that protects from viral infection. Similar to radiation of Toll-like receptors in pathogen immunity, our results establish cGLRs as a diverse family of metazoan pattern recognition receptors.

[Download PDF](#)

Main

To define the function of cGAS-like enzymes in animals, we screened predicted cGAS homologues for suitability in structural analysis and determined a 2.4 Å crystal structure of the human protein MB21D2 (hMB21D2; encoded by *C3orf59*) and a 1.6 Å crystal structure of a protein from the beetle species *Tribolium castaneum* (GenBank XP_969398.1) (Supplementary Table 1). Despite divergence in the primary sequence, the hMB21D2 and *T. castaneum* XP_969398.1 structures each reveal close homology to human cGAS with a shared bi-lobed architecture, a caged nucleotidyltransferase core, a Gly-[Gly/Ser] activation loop and a putative catalytic triad (Fig. 1a, Extended Data Fig. 1). In human cGAS, the primary ligand-binding surface is a long groove on the back of the enzyme formed by the α-helix spine and a Zn-ribbon motif that is essential for recognition of double-stranded DNA^{3,11,12,13,14}. A conserved groove is present in both the hMB21D2 and the *T. castaneum* XP_969398.1 structures (Fig. 1a), but is notably distinguished by the absence of a Zn-ribbon and the insertion of a C-terminal α-helix in hMB21D2 (Fig. 1b). We hypothesized that the remodelling of this groove controls the detection of distinct ligands. The hMB21D2

surface is overall neutral with no obvious capacity to bind nucleic acids, and no enzymatic activity was detected with a panel of potential activating ligands (Extended Data Fig. 1d,e). In contrast to hMB21D2, the surface of *T. castaneum* XP_969398.1 shares highly conserved basic residues with human cGAS (Fig. 1a) and we therefore tested this enzyme with candidate DNA and RNA ligands. We observed that *T. castaneum* XP_969398.1 is activated to synthesize a nucleotide product upon recognition of double-stranded RNA (dsRNA) (Fig. 1c). Despite exhibiting a clear difference in ligand specificity, analysis of all structures in the Protein Data Bank confirmed that *T. castaneum* XP_969398.1 is a close homologue of mammalian cGAS and is distinct from previously characterized RNA sensors including oligoadenylate synthase 1 (ref. 15) (Extended Data Fig. 1f). Together, these results establish the existence of cGLRs in animals and demonstrate that remodelling of a primary ligand-binding surface enables the recognition of divergent molecular patterns.

Fig. 1: Structural remodelling in animal cGLRs enables divergent pattern recognition.

 figure1

a, Crystal structures and surface electrostatics of hMB21D2 and *Tc*-cGLR. Structural comparison with the human cGAS (hcGAS)–DNA complex (Protein Data Bank (PDB): 6CTA)¹⁴ reveals that cGLRs have a conserved architecture with a nucleotidyltransferase signalling core and a shared primary ligand-binding surface (dashed lines). The purple and green boxes indicate cutaways in **b**. **b**, Zoomed-in cutaways highlighting structural insertions and deletions unique to each cGLR. hMB21D2 and *Tc*-cGLR lack the Zn-ribbon motif present in cGAS (left) and hMB21D2 contains a C-terminal α-helix extension that contacts the central ‘spine’ helix (right). Alterations in the predicted ligand-binding surfaces suggest individual cGLRs are remodelled to recognize different molecular patterns. **c**, Thin-layer chromatography analysis and quantification of *Tc*-cGLR reactions in the presence of nucleic acid ligands. *Tc*-cGLR is specifically activated by dsRNA recognition to synthesize a nucleotide (nt) product. Data are relative to maximum activity and represent the mean ± s.e.m. for $n = 3$ independent experiments. Ori, origin; P_i, inorganic phosphate; ss, single-stranded.

Source data

To identify additional cGLRs that respond to dsRNA, we used the *T. castaneum* cGLR (*Tc*-cGLR) sequence to search for predicted cGAS homologues in species related to the model organism *D. melanogaster*. We identified 153 cGLR genes across 42 species in the order Diptera, which cluster into distinct clades designated 1–5 (Fig. 2a, Supplementary Table 2). *Drosophila* encode a remarkable number of cGLR genes, with individual species predicted to have between three and seven enzymes (Extended Data Fig. 2a). In a systematic biochemical screen, we purified and tested 53 recombinant cGLR proteins and identified active enzymes from the species *Lucilia cuprina*, *Drosophila eugracilis*, *Drosophila erecta* and *Drosophila simulans* (Extended Data Figs. 2b–f, 3a). Similar to *Tc*-cGLR, each active Diptera enzyme specifically responded to dsRNA, indicating that cGLR-based recognition of RNA is conserved across diverse insect species (Fig. 2b, Extended Data Fig. 4a).

Fig. 2: *Drosophila* cGLR1 senses long dsRNA.

 **figure2**

a, Phylogeny representing 153 Diptera cGLR genes clustered into clades 1–5 (less than 30% sequence identity between clades). Forty-one of forty-two analysed Diptera species encode enzymes in clade 5, including *D. melanogaster* CG12970 (cGLR1) and CG30424. Enzymes analysed by forward biochemical screen (red dot) and identified as dsRNA-sensing cGLRs (blue circle) are denoted. **b**, Diptera cGLRs identified in the screen are activated to form a nucleotide product by the dsRNA mimic poly I:C. **c**, Mutation to the *Dm*-cGLR1 active site ablates enzymatic activity. Data in **b** and **c** are representative of $n = 3$ independent experiments. WT, wild type. **d**, *Dm*-cGLR1 in vitro activity was monitored in the presence of a panel of dsRNAs and quantified relative to 40 bp dsRNA. Data are the mean \pm s.e.m. of $n = 3$ independent experiments. **e**, Analysis of *Dm*-cGLR1 activity in human cells using mammalian STING and IFN β reporter induction, quantified relative to WT activity. *Dm*-cGLR1 signalling in cells is dependent on stimulation of dsRNA and mutation of the catalytic site, or the predicted ligand-binding residues ablates activity. Data are mean \pm s.e.m. of $n = 3$ technical replicates and representative of $n = 3$ independent experiments. The inset shows a model of the *Tc*-cGLR–dsRNA complex based on the hcGAS–dsDNA structure (PDB: 6CTA)¹⁴ used to predict the *Dm*-cGLR1 ligand-binding residues R23, K42, K52, R241 and K251.

Source data

The *D. simulans* enzyme identified in our screen shares 91% sequence identity with the protein product of the *D. melanogaster* gene CG12970 (GenBank NP_788360.2). Analysis of recombinant *D. melanogaster* CG12970 protein revealed that it also synthesizes a nucleotide product specifically in the presence of dsRNA and we therefore named this gene *cGAS-like Receptor 1* (*Dm-cGLR1*) (Fig. 2c). To understand how dsRNA activates *Drosophila* cGLR1, we analysed the molecular determinants for enzymatic activity in vitro. We observed that *D. simulans* cGLR1 (*Ds-cGLR1*) and *Dm-cGLR1* recognize dsRNAs longer than 30 bp with no preference for 5' RNA phosphorylation (Fig. 2d, Extended Data Fig. 4b,c). Notably, activation of *Ds-cGLR1* and *Dm-cGLR1* requires dsRNA ligands that exceed the length of 21–23-bp RNA molecules commonly produced during RNA interference in *Drosophila*, suggesting specific avoidance of self-recognition^{16,17}. Similar to the formation of condensates observed with human cGAS recognition of dsDNA¹⁸, *Ds-cGLR1* selectively binds to dsRNA and forms a higher-order complex that is dependent on the length of dsRNA (Extended Data Fig. 5). Ectopic expression of *Dm-cGLR1* or *Ds-cGLR1* in human cells demonstrated that cGLR1 activity is sufficient to enable cellular dsRNA sensing and drive activation of a STING-dependent immune response (Fig. 2e, Extended Data Figs. 3f, 4e). *Dm-cGLR1* and *Ds-cGLR1* signalling in cells required dsRNA stimulation, and mutations to the enzyme catalytic site disrupted downstream activation of STING (Fig. 2e, Extended Data Fig. 3f). To understand how *Drosophila* cGLR1 engages dsRNA, we modelled interactions using the *Tc-cGLR* and human cGAS–DNA structures as a template¹⁴ and observed that charge-swap mutations to the conserved basic ligand-binding surface disrupted product synthesis in vitro and STING signalling in cells (Fig. 2e, Extended Data Fig. 3c–f). Together, these data demonstrate that insect cGLRs and human cGAS use a shared mechanism of ligand detection and reveal that *Dm-cGLR1* can function as a foreign RNA sensor.

A role in sensing long dsRNA suggests that the function of *Dm-cGLR1* is to control a downstream immune response in *Drosophila*. In human cells, cGAS synthesizes the nucleotide second messenger 2'3'-cGAMP, which contains a non-canonical 2'–5' phosphodiester linkage that is required for potent activation of immune signalling^{2,3,4,5}. To determine how *Dm-cGLR1* controls cellular signalling, we purified the nucleotide reaction product for direct comparison to 2'3'-cGAMP. The *Dm-cGLR1* product exhibited a C18 chromatography migration profile distinct from 2'3'-cGAMP and all previously known naturally occurring cyclic dinucleotide (CDN) signals (Fig. 3a, Extended Data Fig. 6a). Production of this nucleotide signal was conserved in Diptera with *Ds-cGLR1*, *Lc-cGLR* and *Deu-cGLR* reactions, each synthesizing the same major reaction product (Extended Data Fig. 6a). Using nucleobase-specific labelling and nuclease digestion of the *Dm-cGLR1* product, we observed a 3'–5' linkage connected to an adenine phosphate and a protected 2'–5' linkage connected

to a guanosine phosphate, indicating a mixed-linkage cyclic GMP–AMP species (Fig. 3b). We verified these findings with comparative high-performance liquid chromatography and tandem mass spectrometry profiling against a chemically synthesized standard, and confirmed that the shared Diptera cGLR product is the novel isomer 3'2'-cGAMP (Fig. 3a,b, Extended Data Fig. 6a,b).

Fig. 3: Discovery of 3'2'-cGAMP as a metazoan nucleotide second messenger.

 figure3

a, High- performance liquid chromatography (HPLC) analysis of the *Dm*-cGLR1 reaction (orange) and comparison with synthetic standards (black or dashed lines) demonstrates that *Dm*-cGLR1 synthesizes 3'2'-cGAMP as the major product. A minor *Dm*-cGLR1 reaction product is 2'3'-c-di-AMP (see also Extended Data Fig. 6a). **b**, Thin-layer chromatography analysis of mouse cGAS and *Dm*-cGLR1 reactions labelled with either α -³²P-ATP or α -³²P-GTP (indicated as [α -³²P]NTP) and treated as indicated. Pairwise labelling and nuclease P1 digestion verify that cGAS and *Dm*-cGLR1 synthesize distinct cGAMP isomers with opposite phosphodiester linkage specificities. Representative of $n = 3$ independent experiments. High-resolution mass spectrometry confirms the major Diptera cGLR product as 3'2'-cGAMP (see also Extended Data Fig. 6b). **c**, HPLC quantification of insect cGLR nucleotide products. 3'2'-cGAMP is the dominant product of each identified Diptera cGLR (denoted by a black line), and 2'3'-cGAMP is the dominant product of cGAS and *Tc*-cGLR. Data are the mean quantified product of $n = 3$ independent experiments. **d**, Thermal denaturation assay showing that dSTING selectively recognizes 3'2'-cGAMP (see also Extended Data Fig. 8b,c). Representative of $n = 3$ independent experiments. **e**, Crystal structure of the dSTING–3'2'-cGAMP complex reveals a tightly closed homodimer and an ordered β -strand lid, indicating high-affinity engagement of the endogenous *Drosophila* second messenger 3'2'-cGAMP. **f**, Alignment and conservation of the stem helix and β -strand lid in human and insect STING proteins. Critical ligand-binding residues (blue dot) and adaptations specific to Diptera (red outline) are denoted. **g**,

Superposition of the dSTING–3'2'-cGAMP (blue–orange) complex and the human STING–2'3'-cGAMP (grey–pink) (PDB: 4KSY)⁵ complex reveals that human STING readout of the 2'–5' phosphodiester bond by R232 is absent in dSTING (left). Human STING S162 (grey) contacts the free 3' OH of the guanosine base in 2'3'-cGAMP (pink). dSTING N159 (blue) extends across the ligand-binding pocket to contact the free 3' OH of the adenosine base in 3'2'-cGAMP (orange) (right).

Dm-cGLR1 synthesizes 3'2'-cGAMP in a two-step reaction through production of the linear intermediate pppA[2'–5']pG and uses an opposite nucleobase reaction order compared with human cGAS^{2,3,19} (Extended Data Fig. 7a). We next used mass spectrometry to analyse lysates expressing each recombinant dipteran cGLR from our screen. 3'2'-cGAMP was detected as a product of 15 cGLRs, including enzymes from each subgroup within clade 5 of the Diptera cGLR phylogeny (Extended Data Fig. 6c). cGLRs clustered within clade 5 collectively represent 41 species, suggesting widespread conservation of 3'2'-cGAMP signalling in Diptera. The beetle enzyme *Tc*-cGLR synthesizes 2'3'-cGAMP, supporting that 2'3'-cGAMP is an ancestral signalling molecule in metazoans and that 3'2'-cGAMP signalling is a recent adaptation in flies^{8,20,21} (Fig. 3c, Extended Data Fig. 6a). Insect and mammalian viruses encode 2'3'-cGAMP-specific nucleases named poxins that allow evasion of cGAS–STING immune responses²². 3'2'-cGAMP was protected from cleavage by poxin (Extended Data Fig. 7b–d), indicating that an isomeric switch in the specificity of phosphodiester linkage endows *Drosophila* with a signalling pathway resistant to a major form of viral immune evasion.

Drosophila STING (dSTING) is known to function as a cyclic dinucleotide receptor in vivo^{23,24,25,26}, but an endogenous nucleotide second messenger has not been previously identified. We therefore developed an in vitro thermo-fluor binding assay to analyse dSTING recognition of specific CDNs. dSTING preferentially formed a thermostable complex with 3'2'-cGAMP and exhibited no detectable complex formation with 2'3'-cGAMP or other CDNs in vitro (Fig. 3d, Extended Data Fig. 8b,c). Using direct delivery of CDNs to permeabilized cells, we confirmed that dSTING preferentially responds to 3'2'-cGAMP in the cellular environment (Extended Data Fig. 8d). To define the mechanism of selective 3'2'-cGAMP recognition, we next determined a 2.0 Å crystal structure of the *D. eugracilis* STING (GenBank XP_017066673) CDN-binding domain in complex with 3'2'-cGAMP (Fig. 3e, Supplementary Table 1). dSTING adopts a highly conserved V-shaped homodimeric architecture with a deep central pocket that binds to 3'2'-cGAMP. The dSTING–3'2'-cGAMP structure reveals a tightly ‘closed’ conformation with dSTING protomers positioned 36 Å apart, similar to the closed conformation of human STING bound to 2'3'-cGAMP⁵ (Extended Data Fig. 8e). Each nucleobase of 3'2'-cGAMP is stacked between dSTING Y164 and R234, and E257 specifically coordinates the 3'2'-cGAMP guanosine N2 position (Extended Data Fig. 8f). In human STING, high-affinity

recognition of 2'3'-cGAMP requires readout of the 2'-5' phosphodiester linkage by R232 in the β-strand lid⁵. In dSTING, the equivalent R229 makes no contact with either phosphodiester bond. Instead, R229 is repositioned to extend outwards from the ligand-binding pocket by the deletion of a single lid residue and the formation of a salt bridge with E267 on the opposing protomer, explaining the diminished affinity of dSTING for 2'3'-cGAMP (Fig. 3f,g). In addition, a key asparagine substitution, N159, in dSTING extends across the binding pocket to coordinate the adenosine 3' OH in 3'2'-cGAMP and directly replaces the human STING S162 residue that contacts the guanosine 3' OH in 2'3'-cGAMP (Fig. 3f,g). We tested a panel of dSTING-mutant proteins and confirmed that mutations to each coordinating residue disrupt the formation of the dSTING–3'2'-cGAMP complex (Extended Data Fig. 8i). The unique adaptations in the ligand-binding pocket observed in the dSTING–3'2'-cGAMP structure are widely conserved in Diptera and together explain a mechanism for how specific 3'2'-cGAMP-dependent signalling drives the activation of dSTING.

To determine how *Dm*-cGLR1–3'2'-cGAMP–dSTING signalling controls immune responses in vivo, we next injected 3'2'-cGAMP into *D. melanogaster* to directly monitor the dSTING response. 3'2'-cGAMP potently induced the expression of *Sting* and three other *Sting-regulated genes* (*Srg*) in a dose-dependent manner (Fig. 4a, Extended Data Fig. 9). Notably, 3'2'-cGAMP-dependent signalling through dSTING was significantly more potent than the response triggered by injection of the bacterial CDN signal 3'3'-c-di-GMP (Fig. 4a, Extended Data Fig. 9e–k). Genetic mutations to *Sting* and the NF-κB homologue *Relish* ablated 3'2'-cGAMP-induced responses, demonstrating that signalling operates through a conserved dSTING–NF-κB pathway (Fig. 4a, Extended Data Fig. 9e–k). We challenged flies with viral infection and observed that 3'2'-cGAMP markedly suppressed the replication of two unrelated RNA viruses: *Drosophila* C virus (*Dicistroviridae*), a natural *Drosophila* pathogen, and vesicular stomatitis virus (*Rhabdoviridae*) (Fig. 4b,c, Extended Data Fig. 10a,b). 3'2'-cGAMP activation of antiviral immunity was strictly dependent on *Sting* and resulted in a response that significantly delayed pathogen-mediated mortality (Fig. 4b,c, Extended Data Fig. 10a,b). Direct comparison of the protective effects against *Drosophila* C virus infection showed that the endogenous signal 3'2'-cGAMP exhibited greater antiviral potency than 2'3'-cGAMP. 3'2'-cGAMP more robustly suppressed RNA viral loads and extended animal survival (Fig. 4d, Extended data Fig. 10c,d), revealing that the dSTING antiviral signalling axis is preferentially activated by 3'2'-cGAMP in vivo. Together, these results demonstrate that 3'2'-cGAMP is an antiviral nucleotide second messenger in *D. melanogaster* and establish a cGLR–STING–NF-κB axis that protects animals from viral replication.

Fig. 4: 3'2'-cGAMP activates STING-dependent antiviral immunity in *Drosophila*.

 **figure4**

a, Synthetic 3'2'-cGAMP or 3'3'-c-di-GMP was injected into the body cavity of flies and gene expression was measured after 24 h. *Sting-regulated gene 1 (Srg1)* RNA levels are shown as fold induction compared with buffer control in WT. dSTING^{Mut} is the RXN mutant, as previously characterized^{[23,26](#)}. Data in **a** and **b** are mean ± s.e.m. of

RNA levels measured relative to the control gene *Rpl32* from $n = 3$ independent experiments of $n = 6$ flies. The P values, calculated by unpaired t -test, were not significant (NS) unless otherwise noted; NS $P > 0.05$. **** $P < 0.001$, * $P = 0.0119$. **b**, Viral RNA loads 3 days after infection with *Drosophila* C virus (DCV) demonstrate significantly diminished viral replication in WT flies injected with 3'2'-cGAMP. ** $P = 0.051$. **c**, Survival analysis of animals infected with DCV demonstrates that injection of 3'2'-cGAMP results in a *Sting*-dependent response that significantly delays mortality. Data are mean \pm s.e.m. **** $P < 0.001$. Data in **c** and **d** are each from $n = 3$ independent experiments of $n = 30$ flies. dpi, days post-infection. **d**, Survival analysis directly comparing the effects of cGAMP isomers 7 days after DCV infection. 3'2'-cGAMP injection increases animal survival in a dose-dependent manner and confers greater protection than 2'3'-cGAMP (see also Extended Data Fig. 10d). **e**, Proposed model for cGLR–STING signalling. Upon recognition of distinct molecular patterns, animal cGLRs synthesize a nucleotide second messenger that activates antiviral immunity through STING.

Source data

Along with cGAS recognition of dsDNA, the discovery of animal cGLR dsRNA sensors establishes a diverse class of pattern recognition receptors conserved throughout metazoans. Divergent structural homologues of cGAS in humans and insects demonstrate that cGLRs constitute a rapidly evolving family of proteins in which remodelling of a primary binding surface enables the detection of diverse ligands. Our mechanistic characterization of *Drosophila* cGLR1 activation shows that cGLRs function as direct sensors of pathogen-associated molecular patterns and synthesize distinct second messengers to control a conserved downstream signalling axis (Fig. 4e). *Drosophila* were previously thought to respond to foreign nucleic acid exclusively through RNA interference and direct cleavage of pathogen RNA^{16,17}. *Drosophila* cGLR1 reveals a parallel signalling system for sensing dsRNA and directing an inducible immune response through dSTING. Synthesis of the second messenger 3'2'-cGAMP by *Drosophila* cGLR1 and selective recognition by dSTING demonstrates that metazoans use CDNs beyond 2'3'-cGAMP as endogenous second messengers and highlights the evolutionary plasticity of cGLR signalling. Our structural analysis also reveals that the human cGLR MB21D2 is competent for synthesis of nucleotide second messengers and has a remodelled ligand-binding groove that is probably adapted for detection of an unknown stimulus. Together with the known high frequency of hMB21D2 mutations in cancer^{27,28}, these results support a more extensive role for cGLR signalling in human biology. The existence of multiple unique cGLRs encoded within a single species (Extended Data Fig. 2a) suggests a model in which the cGLR signalling scaffold is harnessed to detect several distinct stimuli. In support of this conclusion, Hartmann, Imler, Cai and colleagues have identified cGLR2 as a second functional cGLR in *Drosophila* and have demonstrated

in vivo that cGLR1 and cGLR2 have discrete roles in *Drosophila* immunity²⁹. Together, our results define cGLRs as receptors in animal cells that are capable of detecting diverse pathogen-associated molecular patterns and dictating response to the foreign environment.

Methods

Bioinformatics and dipteran cGLR sequence analysis

Building on previous analyses^{6,7,8,9,10,30,31}, animal cGAS homologues suitable for crystallography were identified using the amino acid sequences of human cGAS (hcGAS) and *D. melanogaster* CG7194 to seed a position-specific iterative BLAST (PSI-BLAST) search of the NCBI non-redundant protein database. The PSI-BLAST search was performed with an *E* value cut-off of 0.005 for inclusion into the next search round, BLOSUM62 scoring matrix, gap costs settings existence 11 and extension 1, and using conditional compositional score matrix adjustment. Candidate homologues identified from this search included the uncharacterized human protein MB21D2 and the *T. castaneum* sequence XP_969398.1. Pairwise structural comparison between hMB21D2, *Tc*-cGLR and protein structures in the PDB was performed using DALI³², and Z-scores for homologues less than 90% identical to one another (PDB90) were plotted in GraphPad Prism. A Z-score of 15 for *Tc*-cGLR and 13 for hMB21D2 was selected as a lower cut-off to emphasize directly relevant homologues in analysis.

Following structure determination of hMB21D2 and *T. castaneum* XP_969398.1, predicted cGLRs were further identified in Diptera using PSI-BLAST searches seeded with either *D. melanogaster* CG7194 or the *Tc*-cGLR sequence, selecting in each round for proteins matching known cGLR domain organization and active-site residues. Diptera cGLR sequences were aligned using MAFFT (FFT-NS-i iterative refinement method)³³ and used to construct a phylogenetic tree in Geneious Prime v2020.12.23 using the neighbour-joining method and Jukes–Cantor genetic distance model with no outgroup. Further manual analysis and curation of candidate cGLR sequences were performed based on alignments and predictive structural homology using HHpred³⁴ and Phyre2³⁵. Sequences were selected for predicted structural homology to cGAS, including the presence of a conserved nucleotidyltransferase domain with a G[S/G] activation loop and a [E/D]h[E/D] X_{50–90} [E/D] catalytic triad. Manual refinement was also used to exclude duplicate sequences, gene isoforms and proteins less than 250 or greater than 700 residues. NCBI available genomes from 42 species in Diptera are represented in the final tree, including 31 species in the genus *Drosophila*. Clustering of sequences in the final unrooted tree was used to define clades, with no more than 30% sequence identity shared between members of different

clades. Further manual analysis of the tree was used to determine the number and distribution of predicted cGLRs by species (see Extended Data Fig. 2a).

PROMALS3D³⁶ was used for structure guided alignment of apo hcGAS (PDB: 4KM5)¹², hMB21D2 and *Tc*-cGLR in Extended Data Fig. 1a. MAFFT (FFT-NS-i iterative refinement method)³³ was used to align STING sequences in Extended Data Fig. 8a. Geneious Prime software was used to generate the sequence alignments in Fig. 3f and Extended Data Figs. 1a, 3a, 8a.

Protein expression and purification

Recombinant cGLR and dSTING proteins were expressed and purified using methods previously optimized for hcGAS¹⁴. Animal cGLR and dSTING sequences were codon-optimized for expression in *Escherichia coli* and cloned from synthetic constructs (GeneArt or Integrated DNA Technologies) into a custom pET16 expression vector with an N-terminal 6× His–MBP fusion tag or an N-terminal 6× His–SUMO2 fusion. The full-length coding sequence was used except for hcGAS 157–522, mouse cGAS 147–607, hMB21D2 S29–F491, *Ds*-cGLR1 19–393 and *D. eugracilis* STING 150–340 as specified. The N terminus of *D. eugracilis* STING 150–340 was fused to the full-length coding sequence of T4 lysozyme connected by a Gly-Ser linker sequence. Briefly, transformed BL21-CodonPlus(DE3)-RIL *E. coli* (Agilent) were grown in MDG media overnight before inoculation of M9ZB media at an OD₆₀₀ of 0.0475. M9ZB cultures were grown to OD₆₀₀ of 2.5 (approximately 5 h at 37 °C with shaking at 230 rpm) followed by cooling on ice for 20 min. Cultures were induced with 500 μM IPTG before incubation at 16 °C overnight with shaking at 230 rpm. Cultures were pelleted the following day and either flash frozen in liquid nitrogen for storage at –80 °C or directly lysed for purification. Selenomethionine-substituted proteins for crystallography experiments were purified using a modified growth protocol as previously described³⁷.

For large-scale protein purification, proteins were expressed with a 6× His–SUMO2 (*Tc*-cGLR, *Ds*-cGLR1, *Deu*-cGLR, *Lc*-cGLR and dSTING) or 6× His–MBP (*Dm*-cGLR1 and *Der*-cGLR1) fusion tag and grown as approximately 4–8 × 1 l cultures in M9ZB media. Pellets were lysed by sonication in lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 30 mM imidazole, 10% glycerol and 1 mM dithiothreitol) and clarified by centrifugation at approximately 47,850g for 30 min at 4 °C and subsequent filtration through glass wool. Recombinant protein was purified by gravity flow over Ni-NTA resin (Qiagen). Resin was washed with lysis buffer supplemented to 1 M NaCl and then eluted with 20 ml of lysis buffer supplemented to 300 mM imidazole. SUMO2 fusion proteins were cleaved by supplementing elution fractions with approximately 250 μg of human SENP2 protease (D364–L589 with M497A mutation) during overnight dialysis at 4 °C against dialysis buffer (20 mM HEPES pH 7.5, 250 mM KCl and 1 mM dithiothreitol). MBP-tagged fusion proteins were buffer

exchanged into lysis buffer with 4% glycerol and no imidazole to optimize conditions for overnight cleavage by recombinant TEV protease at approximately 10 °C. cGLR proteins were next purified by ion-exchange chromatography using 5 ml HiTrap Heparin HP columns (GE Healthcare) and eluted across a 150–1,000 mM NaCl gradient. Target protein fractions were pooled and further purified by size-exclusion chromatography using a 16/600 Superdex 75 column or 16/600 Superdex 200 column (Cytiva) and storage buffer (20 mM HEPES pH 7.5, 250 mM KCl and 1 mM TCEP). Final proteins were concentrated to approximately 20–30 mg ml⁻¹ and flash frozen with liquid nitrogen and stored at –80 °C for crystallography or supplemented with 10% glycerol before freezing for biochemistry experiments. *Tc*-cGLR and *Ds*-cGLR1 mutant proteins were purified from 1 l M9ZB cultures using Ni-NTA affinity chromatography and overnight dialysis directly into storage buffer (20 mM HEPES pH 7.5, 250 mM KCl, 10% glycerol and 1 mM TCEP) without SUMO2 tag cleavage.

For small-scale protein purification used in the Diptera cGLR screen, recombinant proteins were expressed with a 6× His–MBP fusion tag with the exception of hcGAS, mouse cGAS, *Tc*-cGLR, *Deu*-cGLR, *Lc*-cGLR and *Ds*-cGLR1, which were expressed with a 6× His–SUMO2 fusion tag. Small-scale cultures were grown in 20 ml of M9ZB media, lysed with sonication, and recombinant protein was purified as previously described⁹. Briefly, protein was purified directly from lysates by centrifugation and flow-through over Ni-NTA resin (Qiagen) in 2 ml Mini Spin columns (Epoch Life Sciences). Following elution with elution buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 300 mM imidazole, 10% glycerol and 1 mM dithiothreitol), proteins were buffer exchanged into storage buffer (20 mM HEPES pH 7.5, 250 mM KCl, 10% glycerol and 1 mM TCEP). Fresh protein preparations were immediately used for in vitro nucleotide synthesis reactions.

Protein crystallization and structure determination

Crystals of native and selenomethionine-substituted hMB21D2 S29–F491, *Tc*-cGLR and T4 lysozyme-dSTING L150–I340 were grown at 18 °C using hanging-drop vapour diffusion. Optimized crystals were grown in EasyXtal 15-well trays (NeXtal Biotechnologies) with 350 µl of reservoir solution and 2-µl drops set with a ratio of 1 µl of protein solution and 1 µl of reservoir solution. hMB21D2 crystals were grown using the reservoir solution (1.2 M ammonium sulfate, 5 mM MgCl₂ and 100 mM MES pH 6.2) based on conditions previously identified by Wang and Huang (University of Illinois at Urbana-Champaign)³⁸ for 1 day before cryoprotection with reservoir solution supplemented with 30% glycerol and freezing in liquid nitrogen. *Tc*-cGLR crystals were grown using the reservoir solution (0.3 M potassium thiocyanate and 10–16% PEG-3350) for 5–16 days before cryoprotection with reservoir solution supplemented with 15% ethylene glycol and freezing in liquid nitrogen. Apo T4 lysozyme-dSTING crystals were grown using the reservoir solution (0.2 M sodium

citrate, 0.1 M Tris-HCl and 22% PEG-3350) 7 days before cryoprotection with reservoir solution supplemented with 15% ethylene and freezing in liquid nitrogen. T4 lysozyme-dSTING-3'2'-cGAMP crystals were grown using the reservoir solution (0.1–0.2 M sodium acetate pH 4.8, 0.2 M ammonium formate and 20–22% PEG-3350) supplemented with 250 µM 3'2'-cGAMP (Biolog) for 10 days before cryoprotection with reservoir solution supplemented to 35% PEG-3350 and freezing in liquid nitrogen. Growth of single hMB21D2 and *Tc*-cGLR crystals was further optimized with streak seeding. X-ray diffraction data were collected at the Advanced Photon Source beamlines 24-ID-C and 24-ID-E and at the Advanced Light Source beamlines 5.0.1 and 8.2.2. Data were processed with XDS and Aimless³⁹ using the SSRL autoxds script (A. Gonzales, SSRL, Stanford, CA, USA). Experimental phase information for all proteins was determined using data collected from selenomethionine-substituted crystals. Anomalous sites were identified, and an initial map was generated with AutoSol within PHENIX⁴⁰. Structural modelling was completed in Coot⁴¹ and refined with PHENIX. Final structures were refined to stereochemistry statistics for the Ramachandran plot (favoured/allowed), rotamer outliers and MolProbity score as follows: hMB21D2, 97.72%/2.28%, 0.71% and 1.27; *Tc*-cGLR, 98.17%/1.57%, 0.28% and 1.02; dSTING apo, 98.00%/2.00%, 0.33% and 1.30; and dSTING-3'2'-cGAMP, 97.06%/2.86%, 1.72% and 1.63. See Supplementary Table 1 and the ‘Data availability’ section for deposited PDB codes. All structure figures were generated with PyMOL 2.3.0.

Nucleotide product synthesis analysis

cGLR nucleotide synthesis activity was analysed by thin-layer chromatography (TLC) as previously described⁹. For the Diptera cGLR screen, recombinant protein preparations were incubated in 10 µl reactions containing 0.5 µl α-³²P-labelled NTPs (approximately 0.4 µCi each of ATP, CTP, GTP and UTP), 200 µM unlabelled NTPs, 10 mM MgCl₂ and 1 mM MnCl₂ in a final reaction buffer of 50 mM Tris-HCl pH 7.5, 100 mM KCl and 1 mM TCEP. Reactions were additionally supplemented with approximately 1 µg poly I:C or 5 µM ISD45 dsDNA as indicated. Reactions were incubated at 37 °C overnight and subsequently treated with 1 µl Quick CIP phosphatase (New England Biolabs) for 20 min at 37 °C to remove unreacted phosphate signal. Each reaction was diluted 1:10 in 100 mM sodium acetate pH 5.2, and 0.5 µl was spotted on a 20-cm × 20-cm PEI-cellulose TLC plate. Plates were run with 1.5 M KH₂PO₄ solvent until approximately 2.5 cm from the top of the plate, dried at room temperature and exposed to a phosphor-screen before signal detection with a Typhoon Trio Variable Mode Imager System (GE Healthcare). For all other nucleotide synthesis reactions visualized by TLC, enzymes were tested at 5 µM with 5 µM nucleic acid ligands and either 1 mM MnCl₂ or 10 mM MgCl₂ for insect cGLRs or cGAS, respectively. hMB21D2 activity was tested with 1 mM MnCl₂ and 10 mM

MgCl_2 using the following synthetic innate immune agonists: lipopeptide Pam3CSK4 (Invivogen), *Staphylococcus aureus* lipoteichoic acid (LTA-SA; Invivogen), *Saccharomyces cerevisiae* cell wall preparation (Zymosan; Invivogen), *Bacillus subtilis* peptidoglycan (PGN-BS; Invivogen), synthetic lipid A mimic (CRX-527; Invivogen), *B. subtilis* flagellin (FLA-BS; Invivogen), imidazoquinoline (Imiquimod; Invivogen), CpG oligonucleotide (ODN 2006; Invivogen) and *S. aureus* 23S rRNA oligonucleotide (ORN Sa19; Invivogen). Other than Diptera screen reactions, samples were not diluted in sodium acetate before spotting on PEI-cellulose TLC plates. TLC images were adjusted for contrast using FIJI⁴² and quantified using ImageQuant (8.2.0). Nucleotide product formation was measured according to the ratio of product to total signal for each reaction. For Figs. 1c, 2d and Extended Data Figs. 3c, d, 4b, 5c, relative activity was calculated as the percent conversion for each reaction relative to maximal conversion observed by wild-type enzyme or in the presence of 40-bp dsRNA for insect cGLRs and 45-bp dsDNA for cGAS.

Electrophoretic mobility shift assay

Analysis of in vitro protein–nucleic acid complex formation was conducted as previously described¹⁴. Briefly, 1 μM 40-bp dsRNA or 45-bp dsDNA was incubated with *Ds*-cGLR1 or hcGAS NTase domain (D157–522) at a concentration of 0.5, 1 or 2 μM . Complex formation was performed with the final reaction buffer (20 mM HEPES-NaOH pH 7.8, 75 mM KCl and 1 mM dithiothreitol). Reactions (20 μl) were incubated at 4 °C for 20 min before separation on a 2% agarose gel using 0.5× TB buffer (45 mM Tris and 45 mM boric acid) as a running buffer. The agarose gel was post-stained in 0.5× TB buffer supplemented with 10 $\mu\text{g ml}^{-1}$ ethidium bromide with gentle shaking at 25 °C for 45 min. Complex formation was visualized using a ChemiDoc MP Imaging System (Bio-Rad).

In vitro condensate formation assays

In vitro condensate formation was analysed as previously described with minor modifications^{18,43}. Briefly, *Ds*-cGLR1 was labelled with AlexaFluor-488 (AF488) carboxylic acid (succinimidyl ester) (Thermo Fisher Scientific) according to the manufacturer’s manuals using a molar ratio of 1:10 at 4 °C for 4 h. Excess free dye was removed by dialysis against buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl and 1 mM dithiothreitol) at 4 °C overnight, and AF488-labelled *Ds*-cGLR1 was then further purified on a PD-10 desalting column (GE Healthcare) eluted with storage buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl and 1 mM TCEP). Final AF488-labelled *Ds*-cGLR1 was concentrated to approximately 5 mg ml^{-1} , flash frozen in liquid nitrogen and stored as aliquots at –80 °C. hcGAS and hcGAS NTase domain (D157–F522) proteins were prepared as previously described⁴³.

To induce condensate formation, *Ds*-cGLR1 (10 µM, containing 1 µM AF488-labelled *Ds*-cGLR1) was mixed with various lengths of RNA (10 µM each) in buffer (20 mM Tris-HCl pH 7.5, 1 mg ml⁻¹ BSA and 1 mM TCEP) in the presence of various salt concentrations at 25 °C in a total reaction volume of 20 µl. The details of proteins, nucleic acids and salt concentrations are indicated in the figures. *Ds*-cGLR1–RNA reactions were placed in 384-well non-binding microplates (Greiner Bio-One) and incubated at 25 °C for 30 min before imaging to allow condensates to settle.

Fluorescence microscopy images were acquired at 25 °C using a Leica TCS SP5 X (Leica Microsystems) mounted on an inverted microscope (DMI6000; Leica Microsystems) with an oil immersion ×63/numerical aperture 1.4 objective lens (HCX PL APO; Leica Microsystems). AF488-labelled *Ds*-cGLR1, hcGAS and hcGAS NTase domain proteins were detected with excitation at 488 nm (emission at 500–530 nm). Microscopy images were processed with FIJI⁴² and contrast adjusted with a uniform threshold setup for each enzyme.

Cellular STING signalling assays

Human HEK293T cells were purchased directly from the American Type Culture Collection (ATCC) and were maintained in complete media (DMEM supplemented with penicillin, streptomycin and 10% FBS) at 37 °C. HEK293T cells were validated by the ATCC and were not tested for mycoplasma contamination. For all assays, 4.5 × 10⁴ cells were plated in 96-well plates. STING and cGLR activity assays were performed using the Dual-Luciferase Reporter Assay System (Promega) as previously described¹², with modifications. Lipofectamine-2000 was used to transfect IFNβ-firefly luciferase and TK-Renilla luciferase reporters and 5 ng of pcDNA4–mouse STING or 15 ng of pcDNA4–dSTING hybrid construct (human STING transmembrane domains fused to the *D. eugracilis* STING CDN-binding domain (L150–I340) appended with the human STING C-terminal tail). For cGLR signalling assays, 150 ng of *Drosophila* cGLR1, 30 ng hcGAS with 120 ng empty vector, or 150 ng empty vector were additionally transfected. The native coding sequence was used for each cGLR and STING pcDNA4 plasmid. Twenty-four to thirty hours after transfection, luciferase was measured using a GloMax microplate reader (Promega), and relative IFNβ expression was calculated by normalizing firefly to Renilla readings. For poly I:C stimulation of cGLR activity, cells were transfected with 100 ng poly I:C (6.125–200 ng for titration experiment) 5 h after plasmid transfection. For dSTING signalling assays, a final concentration of 500 pM to 50 µM 2'3'-cGAMP or 3'2'-cGAMP was delivered to cells using a digitonin permeabilization buffer⁴⁴ 10 h before luciferase measurement.

Nucleotide purification and HPLC analysis

Enzymatic synthesis of cGLR nucleotide products for HPLC analysis was performed using 100- μ l reactions containing 10 μ M cGLR enzyme, 200 μ M ATP, 200 μ M GTP, 10 μ g poly I:C, 1 mM MnCl₂ and 50 mM Tris-HCl pH 7.5. Protein storage buffer (20 mM HEPES pH 7.5, 250 mM KCl and 1 mM TCEP) was used as necessary to adjust KCl concentration to approximately 100 mM. Reactions were incubated at 37 °C for 1 h and then nucleotide product was recovered by filtering reactions through a 30-kDa cut-off concentrator (Amicon) to remove protein. Nucleotide products were separated on an Agilent 1200 Infinity Series LC system using a C18 column (Zorbax Bonus-RP 4.6 × 150 mm, 3.5 μ m) at 40 °C. Products were eluted at a flow rate of 1 ml min⁻¹ with a buffer of 50 mM NaH₂PO₄ pH 6.8 supplemented with 3% acetonitrile.

To purify the *Deu*-cGLR product for mass spectrometry analysis, nucleotide synthesis reaction conditions were scaled as previously described for bacterial cGAS/DncV-like nucleotidyltransferase reactions^{9,45}. Briefly, a 10-ml reaction containing 528 nM *Deu*-cGLR enzyme, 125 μ M ATP, 125 μ M GTP, approximately 250 μ g poly I:C, 1 mM MnCl₂, 50 mM Tris-HCl 7.5 and approximately 25 mM KCl was incubated with gentle rotation for 36 h at 37 °C follow by Quick CIP (NEB) treatment for 6 h. The reaction was monitored using a 20 μ l aliquot supplemented with α -³²P-labelled NTPs and to visualize product formation by TLC. Following incubation, the large-scale reaction was filtered through a 10-kDa concentrator (Amicon) and purified by anion-exchange chromatography using a 1-ml Q-sepharose column (Cytiva) washed with water and eluting with a 0–2 M ammonium acetate gradient. Fractions corresponding to the main product 3'2'-cGAMP were differentiated from fractions corresponding to 2'3'-c-di-AMP by HPLC analysis. Product fractions were further purified by size-exclusion chromatography using a Superdex 30 Increase 10/300 GL (Cytiva) with dH₂O as a running buffer. Peak fractions were eluted in 1-ml volumes, pooled and evaporated for storage before mass spectrometry analysis.

Nucleotide mass spectrometry analysis and 3'2'-cGAMP identification

Purified nucleotide product samples were evaporated at 40 °C under a gentle nitrogen stream. The residual pellet was resuspended in 200 μ l HPLC grade water (J.T. Baker), and 40 μ l was then mixed with 40 μ l of water containing 50 ng ml⁻¹ tenofovir as internal standard and transferred to measuring vials.

Experiments for 3'2'-cGAMP identification were performed on an ACQUITY UPLC I-Class/Vion IMS-QTOF high-resolution LC–MS system (Waters Corporation). Reverse-phase chromatographic separation was carried out at 30 °C on a C18 column (Nucleodur Pyramid C18 50 × 3 mm; 3 μ m Macherey Nagel) connected to a C18 security guard (Phenomenex) and a 2- μ m column saver. Separation was achieved using a binary gradient of water containing 10 mM ammonium acetate and 0.1% acetic

acid (solvent A) and methanol (solvent B). The analytes were eluted at a flow rate of 0.6 ml min⁻¹. The eluting programme was as follows: 0–4 min: 0% B, 4–7.3 min: 0–10% B. This composition of 10% B was held for 1 min, then the organic content was increased to 30% within 2.7 min. The column was then re-equilibrated to 0% B for 2 min. The total analysis run time was 13 min. High-resolution mass spectrometry data were collected on a Vion IMS-QTOF mass spectrometer equipped with an electrospray ionization source, operating in positive ionization mode. The capillary voltage was set at 2.5 kV and the cone voltage at 40 V. The source temperature and desolvation gas temperature was 150 °C and 600 °C, respectively. Analyte fragmentation was achieved using argon as the collision gas. Collision energy of 10 V was used to obtain a low collision energy spectrum. For high collision energy spectrum, the collision energy was ramped from 15 to 30 V. Data acquisition was controlled by the UNIFI 1.9.4.0 software (Waters). For 3'2'-cGAMP identification, the retention times, drift times and fragment spectra of a synthetic 3'2'-cGAMP standard (Biolog) were collected as a reference and compared with those of the suspected 3'2'-cGAMP in the samples.

3'2'-cGAMP quantification

For quantification of 3'2'-cGAMP, chromatographic conditions were transferred to a API4000 mass spectrometer (Sciex) coupled to a Shimadzu HPLC system (Shimadzu). The analytes were ionized by means of electrospray ionization in positive mode applying an ion spray voltage of 3,000 V. Further electrospray ionization parameters were as follows: curtain gas (CUR): 30 psi; collision gas (CAD): 9; source temperature: 650 °C; gas 1: 60 psi and gas 2: 45 psi, respectively. Detection was performed in SRM mode, selecting first for the double-protonated parent ion of 3'2'-cGAMP and 3'3'-cGAMP (used in calibrator series). This resulted in the following mass transitions: 3'2'-cGAMP and 3'3'-cGAMP: *m/z* 338.2 → 152 (quantifier), *m/z* 338.2 → 136 (identifier). Tenofovir served as the internal standard (*m/z* 288 → 176).

For 3'2'-cGAMP semiquantitative quantification from lysate samples in the Diptera cGLR screen, calibration curves were created by plotting peak area ratios of 3'3'-cGAMP as an internal standard versus the nominal concentration of the calibrators. The calibration curve was calculated using quadratic regression and 1/x weighting.

Synthetic cyclic dinucleotide standards

Synthetic nucleotide standards used for HPLC analysis and mass spectrometry analysis were purchased from Biolog Life Science Institute: 3'3'-cGAMP (cat no. C 117), 2'3'-cGAMP (cat no. C 161), 3'2'-cGAMP (cat no. C 238), 2'3'-c-di-AMP (cat no. C 187) and 2'3'-c-di-GMP (cat no. C 182).

Nuclease P1 and poxin cleavage analysis

Nuclease P1 cleavage analysis was performed using *Dm*-cGLR1 reactions labelled with either α -³²P-ATP or α -³²P-GTP as previously described^{9,19}. Briefly, radiolabelled nucleotide products were incubated with nuclease P1 (80 mU; N8630, Sigma) in buffer (30 mM NaOAc pH 5.3, 5 mM ZnSO₄ and 50 mM NaCl) for 30 min in the presence of Quick CIP (NEB).

Poxin cleavage reactions were carried out using purified insect viral AcNPV enzyme as previously described^{22,37}. For HPLC analysis of poxin cleavage, 100- μ l reactions were performed using 100 μ M synthetic 2'3'-cGAMP or 3'2'-cGAMP, 50 nM AcNPV poxin, 50 mM HEPES pH 7.5, 10 mM KCl and 1 mM TCEP. Reactions were incubated at 37 °C and at each specified time reactions were terminated by heat inactivation at 95 °C for 2 min before HPLC analysis as described above. For TLC analysis of poxin cleavage, reactions were performed using α -³²P-GTP-labelled 2'3'-cGAMP synthesized by mcGAS or 3'2'-cGAMP synthesized by *Deu*-cGLR in 5- μ l reactions containing 2.5 μ M nucleotide product and 1 μ M AcNPV poxin, 50 mM HEPES pH 7.5, 10 mM KCl and 1 mM TCEP. Reactions were incubated at 37 °C and at each specified time reactions were terminated by heat inactivation at 80 °C for 5 min before PEI-cellulose TLC analysis as described above.

STING CDN thermal shift assay

A final concentration of 15 μ M dSTING was mixed with 3 \times SYPRO orange dye and 100 μ M synthetic CDN (Biolog) (or a 40 nM to 100 μ M concentration gradient as described in Extended Data Fig. 8c) in 20 mM HEPES-KOH pH 7.5 and 100 mM KCl. Samples were heated from 20 to 95 °C in a Bio-Rad CFX thermocycler with HEX channel fluorescence measurements taken every 0.5 °C. The derivative of each curve over time was calculated using GraphPad Prism and graphed as a percent maximum change in fluorescence or used to calculate the melting temperature.

D. melanogaster cyclic dinucleotide injection and signalling analysis

Fly stocks were raised on standard cornmeal agar medium at 25 °C. All fly lines used in this study were *Wolbachia* free. *w*¹¹¹⁸, *dSTING*^{Control} and *dSTING*^{Rxn} stocks have been described previously^{23,26}. *Relish*^{E20} flies isogenized to the DrosDel *w*¹¹¹⁸ isogenic background were a kind gift from L. Teixeira (Instituto Gulbenkian de Ciéncia)⁴⁶. Cyclic dinucleotides including 3'2'-cGAMP (Biolog), 2'3'-cGAMP (Invivogen) and 3'3'-c-di-GMP (Invivogen) were dissolved in 10 mM Tris-HCl pH 7.5 and diluted to the indicated concentrations. Adult flies (3–5-day old) were injected with 69 nl of cyclic dinucleotide solution or 10 mM Tris-HCl pH 7.5 (negative control) by intrathoracic injection using a Nanoject II apparatus (Drummond Scientific). Flies were collected 24 h later in pools of 6 individuals (3 males and 3

females) or 10 individuals (5 males and 5 females) and homogenized for RNA extraction and quantitative PCR with reverse transcription (RT–qPCR) analysis, as described²⁶. The sample size for all *Drosophila* experiments was determined using previously published protocols²⁶. Flies were randomly selected for each experimental group and blinding was not performed.

***D. melanogaster* viral challenge assays**

For 3'2'-cGAMP and virus co-injection, flies were injected with 69 nl of virus (DCV: 5 plaque-forming units (p.f.u.), vesicular stomatitis virus (VSV): 2,000 p.f.u.) in 10 mM Tris-HCl pH 7.5 or in a 0.9 mg ml⁻¹ 3'2'-cGAMP solution. For titration experiments comparing cGAMP isomers, 69 nl of DCV (5 p.f.u.) in serial diluted concentrations of 2'3'-cGAMP or 3'2'-cGAMP were injected in the body cavity of the flies. Survival was monitored daily, and flies were collected in pools of 6 individuals (3 males and 3 females) or 10 individuals (5 males and 5 females) at the indicated time points to monitor the viral RNA load by RT–qPCR.

Statistical analyses

All statistical analyses were performed using GraphPad Prism 9.0.1. Error bars and sample size for each experiment are defined in the figure legends. Comparisons between groups for gene expression and viral loads were analysed by unpaired parametric *t*-test, two-tailed with no corrections; comparison between groups for survival curves following viral infection were analysed by log-rank test.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Coordinates and structure factors of hMB21D2, *Tc*-cGLR, dSTING and the dSTING–3'2'-cGAMP complex have been deposited in the PDB under the accession codes [7LT1](#), [7LT2](#), [7MWY](#) and [7MWZ](#). All other data are available in the paper or the supplementary materials. [Source data](#) are provided with this paper.

References

1. 1.

Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).

2. 2.

Ablasser, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).

3. 3.

Gao, P. et al. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP–AMP synthase. *Cell* **153**, 1094–1107 (2013).

4. 4.

Diner, E. J. et al. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3**, 1355–1361 (2013).

5. 5.

Zhang, X. et al. Cyclic GMP–AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51**, 226–235 (2013).

6. 6.

Wu, J. & Chen, Z. J. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu. Rev. Immunol.* **32**, 461–488 (2014).

7. 7.

Wu, X. et al. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* **42**, 8243–8257 (2014).

8. 8.

Kranzusch, P. J. et al. Ancient origin of cGAS–STING reveals mechanism of universal 2',3' cGAMP signaling. *Mol. Cell* **59**, 891–903 (2015).

9. 9.

Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).

10. 10.
de Oliveira Mann, C. C., Kiefersauer, R., Witte, G. & Hopfner, K. P. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci. Rep.* **6**, 27498 (2016).
11. 11.
Civril, F. et al. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).
12. 12.
Kranzusch, P. J., Lee, A. S., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3**, 1362–1368 (2013).
13. 13.
Zhang, X. et al. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep.* **6**, 421–430 (2014).
14. 14.
Zhou, W. et al. Structure of the human cGAS–DNA complex reveals enhanced control of immune surveillance. *Cell* **174**, 300–311.e11 (2018).
15. 15.
Lohöfener, J. et al. The activation mechanism of 2'-5'-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure* **23**, 851–862 (2015).
16. 16.
Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat. Rev. Immunol.* **19**, 31–44 (2019).
17. 17.
Schneider, J. & Imler, J. L. Sensing and signalling viral infection in *Drosophila*. *Dev. Comp. Immunol.* **117**, 103985 (2021).

18. 18.
Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* **361**, 704–709 (2018).
19. 19.
Kranzusch, P. J. et al. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011–1021 (2014).
20. 20.
Gui, X. et al. Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature* **567**, 262–266 (2019).
21. 21.
Morehouse, B. R. et al. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).
22. 22.
Eaglesham, J. B., Pan, Y., Kupper, T. S. & Kranzusch, P. J. Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS–STING signalling. *Nature* **566**, 259–263 (2019).
23. 23.
Goto, A. et al. The kinase IKK β regulates a STING- and NF- κ B-dependent antiviral response pathway in *Drosophila*. *Immunity* **49**, 225–234.e4 (2018).
24. 24.
Liu, Y. et al. Inflammation-induced, STING-dependent autophagy restricts Zika virus infection in the *Drosophila* brain. *Cell Host Microbe* **24**, 57–68.e3 (2018).
25. 25.
Martin, M., Hiroyasu, A., Guzman, R. M., Roberts, S. A. & Goodman, A. G. Analysis of *Drosophila* STING reveals an evolutionarily conserved antimicrobial function. *Cell Rep.* **23**, 3537–3550.e6 (2018).
26. 26.

Cai, H. et al. 2'3'-cGAMP triggers a STING- and NF-κB-dependent broad antiviral response in *Drosophila*. *Sci. Signal.* **13**, eabc4537 (2020).

27. 27.

Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).

28. 28.

Gracilla, D. E. et al. Overexpression of wild type or a Q311E mutant MB21D2 promotes a pro-oncogenic phenotype in HNSCC. *Mol. Oncol.* **14**, 3065–3082 (2020).

29. 29.

Holleufer, A. et al. Two cGAS-like receptors induce antiviral immunity in *Drosophila*. *Nature* <https://doi.org/10.1038/s41586-021-03800-z> (2021).

30. 30.

Kranzusch, P. J. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr. Opin. Struct. Biol.* **59**, 178–187 (2019).

31. 31.

Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).

32. 32.

Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).

33. 33.

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

34. 34.

Gabler, F. et al. Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinformatics* **72**, e108 (2020).

35. 35.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

36. 36.

Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* **1079**, 263–271 (2014).

37. 37.

Eaglesham, J. B., McCarty, K. L. & Krantzsch, P. J. Structures of diverse poxin cGAMP nucleases reveal a widespread role for cGAS–STING evasion in host–pathogen conflict. *eLife* **9**, e59753 (2020).

38. 38.

Wang, P. *Discovery and Characterization of Novel RNA Repair Systems*. PhD dissertation, University of Illinois at Urbana-Champaign
<http://hdl.handle.net/2142/78582> (2015).

39. 39.

Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

40. 40.

Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).

41. 41.

Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

42. 42.

Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

43. 43.

Zhou, W., Mohr, L., Maciejowski, J. & Kranzusch, P. J. cGAS phase separation inhibits TREX1-mediated DNA degradation and enhances cytosolic DNA sensing. *Mol. Cell* **81**, 739–755.e7 (2021).

44. 44.

Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. *Science* **328**, 1703–1705 (2010).

45. 45.

Lowey, B. et al. CBASS immunity uses CARF-related effectors to sense 3'-5'- and 2'-5'-linked cyclic oligonucleotide signals and protect bacteria from phage infection. *Cell* **182**, 38–49.e17 (2020).

46. 46.

Ferreira, A. G. et al. The Toll-dorsal pathway is required for resistance to viral oral infection in *Drosophila*. *PLoS Pathog.* **10**, e1004507 (2014).

47. 47.

Cohen, D. et al. Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).

48. 48.

Govande, A. A., Duncan-Lowey, B., Eaglesham, J. B., Whiteley, A. T. & Kranzusch, P. J. Molecular basis of CD-NTase nucleotide selection in CBASS anti-phage defense. *Cell Rep.* **35**, 109206 (2021).

49. 49.

Ye, Q. et al. HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. *Mol. Cell* **77**, 709–722.e7 (2020).

50. 50.

Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc. Natl Acad. Sci. USA* **110**, 1652–1657 (2013).

51. 51.

Zhao, Z. et al. Mn²⁺ directly activates cGAS and structural analysis suggests Mn²⁺ induces a noncanonical catalytic synthesis of 2'3'-cGAMP. *Cell Rep.* **32**, 108053 (2020).

52. 52.

Xie, W. et al. Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl Acad. Sci. USA* **116**, 11946–11955 (2019).

Acknowledgements

We are grateful to P. M. Devant, L. Liu, K. Chat, A. Holleufer, R. Hartmann, T. I. V. Ly, and members of the Kranzusch laboratory for helpful comments and discussion; C. de Oliveira Mann for assistance developing the mass spectrometry analysis of insect cGLRs; and M. Burroughs and A. Iyer for assistance with the bioinformatics analysis of cGAS-like enzymes. The work was funded by grants to P.J.K. from the Pew Biomedical Scholars program, the Burroughs Wellcome Fund PATH program, The Richard and Susan Smith Family Foundation, The Mathers Foundation, The Mark Foundation for Cancer Research, a Cancer Research Institute CLIP grant, a V Foundation V Scholar Award, and the Parker Institute for Cancer Immunotherapy; grants to J.-L.I. and H.C. from the Agence Nationale de la Recherche (ANR-17-CE15-0014), the Investissement d’Avenir Programme (ANR-10-LABX-0036 and ANR-11-EQPX-0022), the Institut Universitaire de France, the Chinese National Overseas Expertise Introduction Center for Discipline Innovation (Project ‘111’ (D18010)), the Foreign Experts Program (2020A1414010306) and The Natural Science Foundation (32000662); a grant from the Deutsche Forschungsgemeinschaft (DFG) within the Priority Program SPP1879 and INST 192/524-1 FUGG; and a grant to A.S.Y.L. from the Pew Biomedical Scholars program. W.Z. is supported as a Benacerraf Fellow in Immunology and through a Charles A. King Trust Postdoctoral Fellowship. B.R.M. is supported as a Ruth L. Kirschstein NRSA Postdoctoral Fellow NIH F32GM133063. X-ray data were collected at the Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), and used a Pilatus detector (S10RR029205), an Eiger detector (S10OD021527) and the Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357), and at beamlines 5.0.1 and 8.2.2 of the Advanced Light Source, a US DOE Office of Science User Facility under contract no.

DE-AC02-05CH11231 and supported in part by the ALS-ENABLE program and the NIGMS grant P30 GM124169-01.

Author information

Author notes

1. Wen Zhou

Present address: School of Life Sciences, Southern University of Science and Technology, Guangdong, China

Affiliations

1. Department of Microbiology, Harvard Medical School, Boston, MA, USA

Kailey M. Slavik, Benjamin R. Morehouse, Adelyn E. Ragucci, Wen Zhou & Philip J. Kranzusch

2. Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA

Kailey M. Slavik, Benjamin R. Morehouse, Adelyn E. Ragucci, Wen Zhou, Amy S. Y. Lee & Philip J. Kranzusch

3. Sino-French Hoffmann Institute, State Key Laboratory of Respiratory Disease, School of Basic Medical Science, Guangzhou Medical University, Guangzhou, China

Xianlong Ai, Yuqiang Chen, Lihua Li, Ziming Wei, Hua Cai & Jean-Luc Imler

4. Research Core Unit Metabolomics, Hannover Medical School, Hannover, Germany

Heike Bähre, Martin König & Roland Seifert

5. Hannover Medical School, Institute of Pharmacology, Hannover, Germany

Roland Seifert

6. Department of Cell Biology, Harvard Medical School, Boston, MA, USA

Amy S. Y. Lee

7. Université de Strasbourg, CNRS UPR9022, Strasbourg, France

Jean-Luc Imler

8. Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute,
Boston, MA, USA

Philip J. Kranzusch

Contributions

Experiments were designed and conceived by K.M.S. and P.J.K. Gene identification and phylogenetic analyses were performed by K.M.S., B.R.M. and P.J.K. *Tc*-cGLR structural experiments were performed by K.M.S. hMB21D2 structural experiments were performed by B.R.M. dSTING structural experiments were performed by K.M.S. and A.E.R. hMB21D2, cGLR and dSTING biochemical experiments were performed by K.M.S. and A.E.R. Phase separation and RNA-binding analysis were performed by W.Z. and A.E.R. Cell biology experiments were designed by K.M.S. and A.S.Y.L., and performed by K.M.S. Nucleotide purification and mass spectrometry experiments were performed by K.M.S., H.B., M.K. and R.S. In vivo *Drosophila* experiments were designed and performed by X.A., Y.C., L.L., Z.W., H.C. and J.-L.I. The manuscript was written by K.M.S. and P.J.K. All authors contributed to editing the manuscript and support the conclusions.

Corresponding author

Correspondence to [Philip J. Kranzusch](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Osamu Nureki 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 Sequence and structural analysis of hMB21D2 and *Tc*-cGLR.

a, Structure guided sequence alignment of the catalytic domain of hcGAS (PDB: 4KM5)¹², hMB21D2 and *Tc*-cGLR. Strict secondary structure conservation further supports conserved structural homology despite primary sequence divergence. The [D/E]hD[X_{50–90}]D catalytic triad is highlighted with a red outline and the human Zn-ribbon insertion that is absent in other cGLRs is denoted with a red dashed outline. hMB21D2 contains an additional 61 residues that are not resolved in the crystal structure and are absent from the alignment. **b**, **c**, Zoomed-in cutaways of the hMB21D2 (**b**) and *Tc*-cGLR (**c**) crystal structures highlighting positioning of conserved catalytic residues in the nucleotidyltransferase active site. In hcGAS, the analogous residues coordinate two Mg²⁺ metal ions to control synthesis of 2'3'-cGAMP (inset, middle; PDB: 6CTA)¹⁴. The hMB21D2 structure is in an inactive state distinguished by misaligned catalytic residues and occlusion of the active site by an extended Gly-Gly activation loop, indicating that catalytic activation is probably controlled by a conformational rearrangement. **d**, **e**, TLC analysis of in vitro tests for potential activating ligands of hMB21D2. No nucleotide products were identified upon stimulation with 40-nt or 40-bp nucleic acid ligands (**d**) or ligands known to activate mammalian Toll-like receptors (**e**). Data shown are representative of $n = 3$ independent experiments. **f**, Z-score structural similarity plot showing homology between hMB21D2 and *Tc*-cGLR with representative structures in the PDB (PDB90). Increasing Z-score indicates greater homology, confirming the close relationship between animal cGLR enzymes and mammalian cGAS, and more distant similarity to cGAS/DncV-like nucleotidyltransferases (CD-NTases) in bacterial antiphage defence systems and human oligoadenylate synthase 1 (refs. 9,15,45,47,48,49,50). Z-score cut-offs are 13 and 15 for hMB21D2 and *Tc*-cGLR, respectively.

Extended Data Fig. 2 Forward biochemical screen of predicted cGLRs in Diptera.

a, Violin plot showing the number of predicted cGLRs in Diptera genomes. *Drosophila* genomes ($n = 31$ species) have a median of four predicted cGLRs in contrast to a median of two predicted cGLRs in other dipteran insects ($n = 11$ species). **b**, Schematic of the in vitro screen of predicted cGLRs in the order Diptera. Fifty-three sequences were selected representing each clade in the phylogeny in Fig. 2a.

Following recombinant protein expression in *E. coli*, lysates were split into two samples for parallel TLC analysis of in vitro enzymatic activity and HPLC-MS analysis of lysate nucleotide metabolites. **c, d**, Purified cGLR proteins were incubated overnight at 37 °C with $\alpha^{32}\text{P}$ -radiolabelled nucleotides, a mixture of Mn²⁺ and Mg²⁺, and the 45-bp immunostimulatory DNA ISD45 or the synthetic dsRNA analogue poly I:C as potential nucleic acid ligands, and reactions were visualized by PEI-cellulose TLC. Wild-type (WT) and catalytically inactive mouse cGAS enzymes were used as controls for each sample set. Note that mouse cGAS exhibits dsDNA-independent activity in the presence of Mn²⁺ (ref. [51](#)). Predicted Diptera cGLRs are grouped by clade (DC01–05) and numbered within each clade. Ligand-dependent activity was identified for DC02_01, DC05_03, DC05_19 and DC05_21; species listed below. We observed ligand-independent activity for two enzymes in clade 3. Data represent $n = 2$ independent experiments. **e**, SDS-PAGE and Coomassie stain analysis of NiNTA-purified cGLR protein fractions used for the biochemical screen. **f**, SDS-PAGE and Coomassie stain analysis of final cGLR proteins used for biochemical studies, which were purified by NiNTA-affinity, ion-exchange chromatography and size-exclusion chromatography.

Extended Data Fig. 3 Sequence analysis and mutagenesis of insect cGLRs.

a, Alignment of the catalytic domain of hcGAS and active cGLRs identified in *T. castaneum*, *D. eugracilis*, *L. cuprina*, *D. erecta*, *D. simulans* and *D. melanogaster*. The EhD[X_{50–90}]D catalytic triad is highlighted with a red outline and the human Zn-ribbon insertion that is absent in insect cGLRs is denoted with a red dashed outline. Predicted basic ligand-binding residues selected for mutational analysis denoted by black circles. cGLRs from *D. erecta* and *D. simulans* are close homologues of *Dm*-cGLR1 (76% and 91% sequence identity, respectively) and thus are also referred to as ‘cGLR1’. All biochemical experiments with *Ds*-cGLR1 were performed with a construct beginning at M19. **b**, In vitro reactions demonstrating that mutation of the catalytic residues ablates nucleotide product synthesis by *Ds*-cGLR1 in response to poly I:C. **c, d**, In vitro reactions analysing dsRNA recognition through the putative ligand-binding surface by *Ds*-cGLR1 (**c**) or *Tc*-cGLR (**d**). The insets for panels **c** and **d** show models of the *Tc*-cGLR–dsRNA complex based on the hcGAS–dsDNA structure (PDB: 6CTA)^{[14](#)}, indicating predicted dsRNA-interacting residues in *Ds*-cGLR1 (**c**) or *Tc*-cGLR (**d**). Charge swap mutation to these residues variably disrupted poly I:C-stimulated activity by *Ds*-cGLR1 and *Tc*-cGLR, shown by TLC (left) and quantified relative to WT activity (right). Data in **b–d** are representative of $n = 3$ independent experiments. **e**, SDS-PAGE and Coomassie stain analysis of purified WT and mutant proteins, as labelled in the above TLC images. **f**, IFN β luciferase assay in which cGLRs are expressed in human cells and CDN synthesis is detected by mammalian STING activation, as in Fig. [2e](#). IFN β was quantified relative to the empty vector

control. In comparison to hcGAS control, which is activated by expression vector-plasmid DNA, *Dm*-cGLR1 (left) and *Ds*-cGLR1 (right) strictly require poly I:C stimulation to activate a downstream STING response. Mutation to catalytic residues or putative ligand-binding residues ablates cGLR1 signalling. See Fig. 2e: *Dm*-cGLR1 activity quantified relative to WT activity upon poly I:C stimulation. Data are mean \pm s.e.m. of $n = 3$ technical replicates and representative of $n = 3$ independent experiments.

[Source data](#)

Extended Data Fig. 4 Analysis of RNA recognition by insect cGLRs.

a–c, In vitro activity assays for each active insect cGLR demonstrating that dsRNA recognition is required for enzyme activation. Reactions were performed with 40-nt or 40-bp synthetic ligands. Weak *Deu*-cGLR ssRNA-stimulated activity may be explained by transient short duplex formation similar to observations that some ssDNA oligos can stimulate mouse cGAS dsDNA-dependent activity³. **b**, TLC and quantification for enzyme activation in the presence of a panel of 10–40-bp synthetic dsRNA ligands. dsRNA (30 bp) is sufficient to stimulate maximal activity for *Tc*-cGLRs, *Dm*-cGLRs and *Lc*-cGLRs, while *Ds*-cGLR1 requires 35 bp and *Deu*-cGLR can be activated by dsRNAs as short as 15 bp. Data are mean \pm s.e.m., quantified relative to maximum observed activity. **c**, Reactions with 146-bp in vitro-transcribed dsRNAs containing either a 5' triphosphate or 5' OH termini demonstrate that dsRNA recognition by insect cGLRs does not involve 5'-end discrimination. **d**, Deconvolution of catalytic metal requirements for enzymatic activity by insect cGLRs. Insect cGLRs require Mn²⁺ for maximal catalytic activity, with weak product formation observed in the presence of Mg²⁺. **e**, Poly I:C titration demonstrates that dsRNA stimulation of *Drosophila* cGLR1 activity in cells is dependent on RNA concentration. IFN β luciferase assay in which cGLRs are expressed in human cells and CDN synthesis is measured by mammalian STING activation, as in Fig. 2e and Extended Data Fig. 3f. IFN β quantified relative to the empty vector control. Data are mean \pm s.e.m. of $n = 3$ technical replicates. All data in a–e represent $n = 3$ independent experiments.

[Source data](#)

Extended Data Fig. 5 Characterization of *Ds*-cGLR1–dsRNA condensate formation.

a, Electrophoretic mobility shift assay (EMSA) showing binding between *Ds*-cGLR1 or the C-terminal NTase domain of hcGAS (hcGAS-NTase) and a 40-bp dsRNA or 45-bp dsDNA. *Ds*-cGLR1 preferentially binds to dsRNA and more weakly interacts with dsDNA, consistent with observed binding between hcGAS and dsRNA¹¹. **b**, EMSA

comparison of *Ds*-cGLR1–dsRNA binding and mammalian cGAS–dsDNA binding. Similar to hcGAS, *Ds*-cGLR1 forms a higher-order protein–nucleic acid complex that does not migrate through the gel, in contrast to the 2:2 binding observed between mouse cGAS and dsDNA. Data in **a** and **b** are representative of $n = 3$ independent experiments. **c**, Analysis of the effect of AF488 labelling on *Ds*-cGLR1 enzymatic activity. Similar to previous observations with hcGAS⁴³, AF488 labelling negatively impacts enzymatic activity but has minimal effect at the ratio of 90% unlabelled and 10% labelled protein used for all imaging experiments. Data are mean \pm s.e.m. of $n = 3$ independent experiments. **d**, **e**, Analysis of hcGAS (**d**) and *Ds*-cGLR1 (**e**) phase separation with AF488-labelled protein. Mammalian cGAS contains a highly disordered N-terminal extension of approximately 150 residues, but this unstructured extension is absent in insect cGLR sequences. In the presence of dsDNA, full-length hcGAS forms highly dynamic liquid droplets^{18,43,52}, whereas the minimal hcGAS NTase domain forms rigid protein–DNA condensates similar to those formed by *Ds*-cGLR1–RNA complexes. hcGAS exhibits a preference for condensate formation in the presence of dsDNA (**d**), whereas *Ds*-cGLR1 exhibits a preference for dsRNA (**e**), as observed in panel **a**. Scale bars, 10 μm . Analysis of *Ds*-cGLR1 dsRNA length specificity for condensate formation demonstrates clear length dependency (**e**) and supports that long dsRNA and condensate formation are required for maximal *Ds*-cGLR1 activation.

Source data

Extended Data Fig. 6 Synthesis of 3'2'-cGAMP by Diptera cGLRs.

a, HPLC analysis of the nucleotide products of *Tc*-cGLR, *Dm*-cGLR1, *Ds*-cGLR1, *Lc*-cGLR and *Deu*-cGLR reactions compared with relevant synthetic controls. Integration of major and minor product peaks in $n = 3$ independent experiments was used to calculate relative product ratios shown in Fig. 3c. **b**, The *Drosophila* cGLR major reaction product was purified from *Deu*-cGLR reactions and compared with synthetic 3'2'-cGAMP with tandem mass spectrometry analysis. Parent mass extracted ion trace (left) and tandem mass spectra comparison (right) validate the chemical identity of the *Drosophila* cGLR product as 3'2'-cGAMP. **c**, Identification of widespread 3'2'-cGAMP synthesis by Diptera cGLRs. The heat map shows the relative concentrations of cGAMP isomers detected by HPLC-MS in bacterial lysates expressing Diptera cGLRs (as described in Extended Data Fig. 2b) (left). In all cases, 3'2'-cGAMP was present as the dominant product with trace amounts of 3'3'-cGAMP and 2'3'-cGAMP detected in some samples as minor species. Right, inset of clade 5 in the Diptera cGLR phylogeny from Fig. 2a annotated to show all enzymes identified to synthesize 3'2'-cGAMP.

Extended Data Fig. 7 Mechanism of 3'2'-cGAMP bond formation and resistance to degradation by viral poxin enzymes.

a, Analysis of *Dm*-cGLR1 reactions with pairwise combinations of α -³²P-labelled nucleotides and non-hydrolyzable nucleotides reveals reaction intermediates and identifies the order of bond formation during 3'2'-cGAMP synthesis. Left: TLC analysis demonstrates that *Dm*-cGLR1 forms a linear intermediate in the presence of GTP and non-hydrolyzable ATP (Apcpp), indicating that the 2'-5' phosphodiester bond is synthesized first. Exposed γ -phosphates removed by phosphatase treatment before analysis are indicated in parentheses. Note that while a linear intermediate cannot be formed in the presence of non-hydrolyzable GTP (Gpcpp), *Dm*-cGLR1 will synthesize the off-product 2'3'-c-di-AMP. Mouse cGAS, which synthesizes 2'3'-cGAMP through the linear intermediate pppG[2'-5']pA, is shown here for comparison¹⁹. Right: schematic of the reaction mechanism for each enzyme. **b**, Poxins are 2'3'-cGAMP-specific viral nucleases that disrupt cGAS-STING signalling. HPLC analysis of synthetic 2'3'-cGAMP or 3'2'-cGAMP treated with poxin from the insect baculovirus *Autographa californica* nucleopolyhedrovirus (AcNPV) is shown^{22,37}. In 1 min, AcNPV poxin cleaves 2'3'-cGAMP into a mixture of intermediate and full-cleavage product; and after 1 h, turnover is complete. No cleavage of 3'2'-cGAMP is observed by AcNPV poxin under these reaction conditions. **c**, Using TLC as a more sensitive assay, we observed minimal cleavage of 3'2'-cGAMP following overnight incubation with AcNPV poxin. Data in **a–c** are representative of $n = 3$ independent experiments. **d**, Schematic highlighting how an isomeric switch in phosphodiester linkage specificity makes 3'2'-cGAMP remarkably resistant to poxin-mediated cleavage.

Extended Data Fig. 8 Structural and biochemical analysis of dSTING.

a, Alignment of the C-terminal CDN-binding domains of human STING, mouse STING, *D. eugracilis* STING and *D. melanogaster* STING. Architecture of the core CDN-binding domain is conserved across metazoans; the disordered C-terminal tail, which controls IRF3–IFN β signalling, is specific to vertebrates^{8,21}. Ligand-interacting residues selected for mutational analysis are denoted with a black circle; Diptera-specific adaptations are highlighted with a red outline. All structural and biochemical experiments were performed with a *D. eugracilis* STING construct terminating at I340. **b**, In vitro thermal denaturation assay analysing dSTING interactions with a panel of CDNs. Only 3'2'-cGAMP forms a thermostable complex with dSTING in vitro (see also Fig. 3d). 2'3'-cGAMP is known to be capable of stimulating dSTING-dependent signalling in vivo²⁶, supporting that dSTING can engage with 2'3'-cGAMP with lower affinity. This observation is consistent with the weaker recognition of bacteria-derived 3'3'-cGAMP and 3'3'-c-di-GMP by human STING^{2,4}. **c**, In vitro

thermal denaturation assay demonstrating concentration-dependent thermal shift induced by 3'2'-cGAMP. **d**, Dose titration of 2'3'-cGAMP and 3'2'-cGAMP in human cells demonstrating selective response by dSTING to 3'2'-cGAMP. The *D. eugracilis* CDN-binding domain (CBD) was adapted for downstream signalling in human cells by addition of N-terminal human transmembrane (hTM) domains and the human C-terminal tail (hCTT). **e**, Comparison of the human STING–2'3'-cGAMP and dSTING–3'2'-cGAMP crystal structures reveals a conserved closed homodimer architecture in which apical ‘wings’ are spread 32–36 Å, demonstrating high-affinity engagement with an endogenous ligand. **f**, Enlarged cutaways of 3'2'-cGAMP in the dSTING crystal structure. Above: the simulated annealing $F_O - F_C$ omit map (contoured at 3 σ). Below: a top-down view highlighting key dSTING–3'2'-cGAMP contacts. **g**, Full crystal structure used to determine the structure of *D. eugracilis* STING in complex with 3'2'-cGAMP. T4 lysozyme is fused to the N terminus of the *D. eugracilis* STING CBD. **h**, Thermal denaturation assay as in Fig. [3d](#) demonstrating that N-terminal fusion of T4 lysozyme does not impair dSTING recognition of 3'2'-cGAMP. **i**, Mutational analysis of key ligand-interacting residues in dSTING; the thermal denaturation assay was used to analyse 3'2'-cGAMP recognition. Mutations that conserve functional contacts with 3'2'-cGAMP (Y164F) maintain ligand recognition; mutations that ablate contacts abrogate ligand binding. N159S exhibits diminished ability to recognize 3'2'-cGAMP. Data in **b** and **i** are mean \pm s.e.m. of the average T_m calculated from $n = 2$ technical replicates in $n = 3$ independent experiments. Data in **c** are representative of $n = 3$ independent experiments. Data in **d** are mean \pm s.e.m. of $n = 3$ technical replicates and representative of $n = 3$ independent experiments. **j**, SDS-PAGE and Coomassie stain analysis of purified WT and mutant proteins.

Source data

Extended Data Fig. 9 3'2'-cGAMP induces the expression of dSTING-regulated genes.

a–d, Injection of 3'2'-cGAMP into *D. melanogaster* has a dose-dependent effect on the expression of *Sting-regulated genes* (*srgs*). 2'3'-cGAMP was used as positive control as previously characterized^{[23,26](#)}. Synthetic nucleotide was injected into the body cavity of WT (w^{1118}) flies and gene expression was measured after 24 h. RNA levels were measured relative to the control gene *RpL32*, and nucleotide concentrations are displayed in $\mu\text{g } \mu\text{l}^{-1}$. Note that for *srg2* measurement after injection of $9\text{E}-7 \mu\text{g } \mu\text{l}^{-1}$ 3'2'-cGAMP, there was one outlier replicate with a value of 0.5977 (data not shown, included in mean analysis). **e–k**, As in Fig. [4a](#), RNA expression analysis of *Sting-regulated genes* (*srgs*) 24 h after injection with synthetic 3'2'-cGAMP or 3'3'-c-di-GMP. RNA levels are shown as fold induction compared with buffer control in WT flies. dSTING^{Mut} = RXN mutant; Relish^{Mut} = Relish^{E20} mutant, as previously

characterized^{23,26}. All data in **a–k** represent the mean \pm s.e.m. of $n = 3$ independent experiments and each point represents a pool of 6 flies. P value ns (>0.05) unless otherwise noted: **** $P < 0.0001$ (**e**); *** $P = 0.0006$, * $P = 0.0404$ (**f**); **** $P < 0.0001$, *** $P = 0.0002$ (**g**); **** $P < 0.0001$ (**h**); ** $P = 0.0015$, * $P = 0.0117$ (**i**); *** $P = 0.0002$, ** $P = 0.0076$ (**j**); *** $P = 0.0009$ (**k**).

[Source data](#)

[**Extended Data Fig. 10 3'2'-cGAMP functions as a potent antiviral ligand.**](#)

a, Analysis of the effect of 3'2'-cGAMP on *Drosophila* C virus (DCV) viral RNA load in flies. dSTING WT and mutant flies were co-injected with DCV and 3'2'-cGAMP or buffer control. Viral RNA levels were measured at each time as indicated relative to the control gene *RpL32*. DCV is a picornavirus-like (+)ssRNA virus in the family *Dicistroviridae*. ** $P = 0.0051$, * $P = 0.0388$. **b**, Analysis of the effect of 3'2'-cGAMP on vesicular stomatitis virus (VSV) viral RNA load in flies. dSTING WT and mutant flies were co-injected with VSV and 3'2'-cGAMP or buffer control as in **a**. Viral RNA levels were measured 4 days post-infection (dpi) relative to the control gene *RpL32*. VSV is a (-)ssRNA virus in the *Rhabdoviridae* family. * $P = 0.0185$. **c**, Analysis of DCV viral RNA load in flies injected with increasing doses of 3'2'-cGAMP, 2'3'-cGAMP or buffer control (as in **a**). Viral RNA levels were measured 2 dpi relative to the control gene *RpL32*. For 2'3'-cGAMP injection: 9E-1 * $P = 0.0192$. For 3'2'-cGAMP injection: 9E-3 * $P = 0.0212$, 9E-2 ** $P = 0.0075$, 9E-1 ** $P = 0.0070$. **d**, Survival curves after DCV infection showing the effect of injection with dose titration of 3'2'-cGAMP or 2'3'-cGAMP compared with buffer control. Both cGAMP isomers significantly delay mortality in a dose-dependent manner; 3'2'-cGAMP provides greater protection in comparison to 2'3'-cGAMP. For 2'3'-cGAMP injection: 9E-3 ** $P = 0.0047$, 9E-2 ** $P = 0.0031$, 9E-1 *** $P = 0.0002$. For 3'2'-cGAMP injection: 9E-4 * $P = 0.0344$, 9E-3 *** $P = 0.0005$, 9E-2 **** $P < 0.0001$, 9E-1 **** $P < 0.0001$. All data in **a–d** represent the mean \pm s.e.m. of $n = 3$ independent experiments and each point represents a pool of 6 flies (**a, b**) or 10 flies (**c, d**). P value is ns unless otherwise noted; ns $P > 0.05$.

[Source data](#)

Supplementary information

[**Reporting Summary**](#)

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

Supplementary Table 1

This table contains a summary of X-ray data collection, phasing and refinement statistics

Supplementary Table 2

This table contains a summary of *Diptera* cGLR sequence information

Source data

[**Source Data Fig. 1**](#)

[**Source Data Fig. 2**](#)

[**Source Data Fig. 4**](#)

[**Source Data Extended Data Fig. 3**](#)

[**Source Data Extended Data Fig. 4**](#)

[**Source Data Extended Data Fig. 5**](#)

[**Source Data Extended Data Fig. 8**](#)

[**Source Data Extended Data Fig. 9**](#)

[**Source Data Extended Data Fig. 10**](#)

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

Slavik, K.M., Morehouse, B.R., Ragucci, A.E. *et al.* cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in *Drosophila*. *Nature* **597**, 109–113 (2021).
<https://doi.org/10.1038/s41586-021-03743-5>

- Received: 19 February 2021
- Accepted: 21 June 2021
- Published: 14 July 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03743-5>

[cGAS-like receptors put a sting into the evolution of immune defence](#)

- Cara West
- Neal Silverman

News & Views 09 Aug 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03743-5>

- Article
- [Published: 14 July 2021](#)

Two cGAS-like receptors induce antiviral immunity in *Drosophila*

- [Andreas Holleufel^{1 na1}](#),
- [Kasper Grønbjerg Winther](#) ORCID: [orcid.org/0000-0001-6146-6603^{2 na1}](#),
- [Hans Henrik Gad](#) ORCID: [orcid.org/0000-0001-8449-1115^{1 na1}](#),
- [Xianlong Ai³](#),
- [Yuqiang Chen³](#),
- [Lihua Li³](#),
- [Ziming Wei³](#),
- [Huimin Deng³](#),
- [Jiyong Liu³](#),
- [Ninna Ahlmann Frederiksen¹](#),
- [Bine Simonsen¹](#),
- [Line Lykke Andersen⁴](#),
- [Karin Kleigrewe](#) ORCID: [orcid.org/0000-0002-7718-0252⁵](#),
- [Louise Dalskov¹](#),
- [Andreas Pichlmair](#) ORCID: [orcid.org/0000-0002-0166-1367^{4,6}](#),
- [Hua Cai](#) ORCID: [orcid.org/0000-0002-6697-652X³](#),
- [Jean-Luc Imler](#) ORCID: [orcid.org/0000-0003-0740-8319^{2,3}](#) &
- [Rune Hartmann](#) ORCID: [orcid.org/0000-0003-1159-066X¹](#)

[Nature](#) volume 597, pages 114–118 (2021)

- 7564 Accesses
- 47 Altmetric

- [Metrics details](#)

Subjects

- [Infection](#)
- [Pattern recognition receptors](#)
- [Virus–host interactions](#)

Abstract

In mammals, cyclic GMP–AMP (cGAMP) synthase (cGAS) produces the cyclic dinucleotide 2'3'-cGAMP in response to cytosolic DNA and this triggers an antiviral immune response. cGAS belongs to a large family of cGAS/DncV-like nucleotidyltransferases that is present in both prokaryotes¹ and eukaryotes^{2,3,4,5}. In bacteria, these enzymes synthesize a range of cyclic oligonucleotides and have recently emerged as important regulators of phage infections^{6,7,8}. Here we identify two cGAS-like receptors (cGLRs) in the insect *Drosophila melanogaster*. We show that cGLR1 and cGLR2 activate Sting- and NF-κB-dependent antiviral immunity in response to infection with RNA or DNA viruses. cGLR1 is activated by double-stranded RNA to produce the cyclic dinucleotide 3'2'-cGAMP, whereas cGLR2 produces a combination of 2'3'-cGAMP and 3'2'-cGAMP in response to an as-yet-unidentified stimulus. Our data establish cGAS as the founding member of a family of receptors that sense different types of nucleic acids and trigger immunity through the production of cyclic dinucleotides beyond 2'3'-cGAMP.

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: cGLR1 and cGLR2 activate the Sting pathway and protect against viral infection.

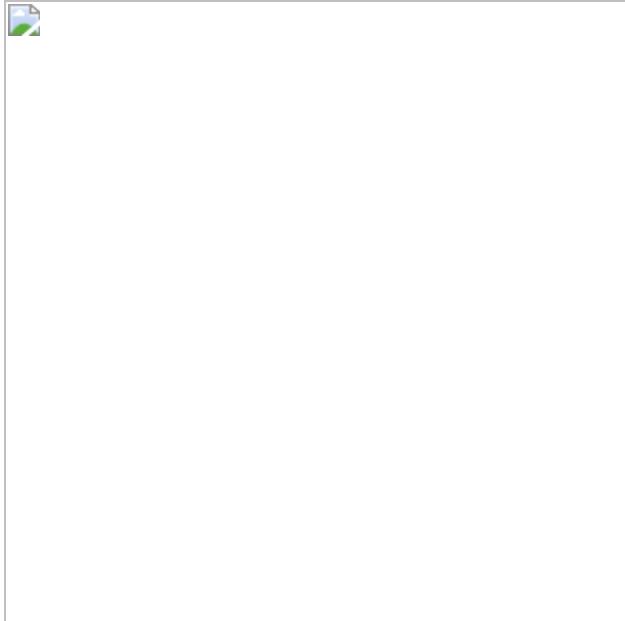


Fig. 2: Loss of cGLR1 or cGLR2 leads to an impaired antiviral immune response in vivo.

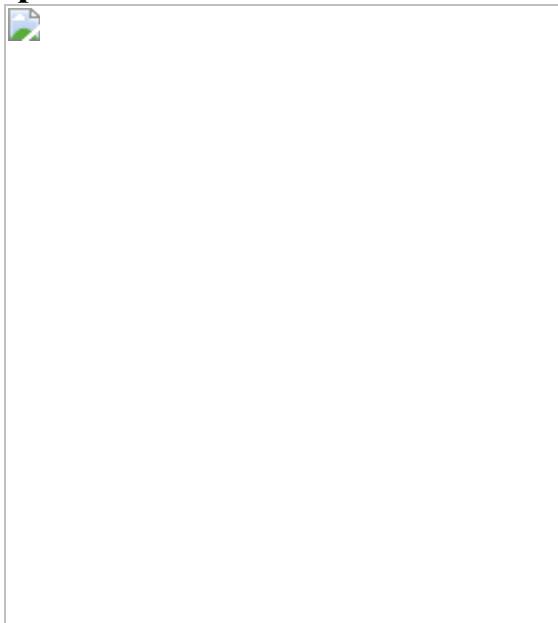
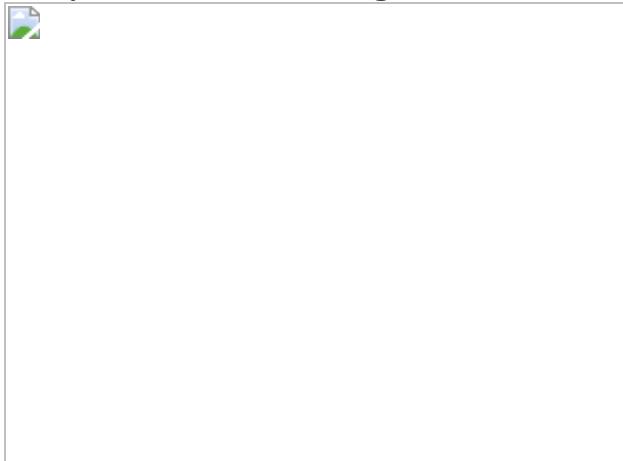


Fig. 3: cGLR1 is activated by dsRNA, whereas cGLR2 is activated by an as-yet-unidentified ligand.



Data availability

The authors declare that the data supporting the findings of this study are available within the Article and its [Supplementary Information](#). The sequences and structures used in this study are CG12970 (UniProt: A1ZA55), CG30424 (UniProt: A8DYP7), CG4746 (UniProt: Q9U3W6), CG4766 (UniProt: Q9Y106), CG7194 (UniProt: Q9VSH0), porcine OAS1 (UniProt: Q29599), human OAS1 (UniProt: P00973-1), human cGAS

(UniProt: Q8N884-1), porcine cGAS (UniProt: I3LM39), mouse cGAS (UniProt: Q8BSY1) and mouse cGAS in complex with DNA and a cGAMP intermediate analogue (Protein Data Bank code 4K98). [Source data](#) are provided with this paper.

References

1. 1.
Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
2. 2.
Ablasser, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).
3. 3.
Civril, F. et al. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).
4. 4.
Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
5. 5.
Wu, J. et al. Cyclic GMP–AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
6. 6.
Lowey, B. et al. CBASS immunity uses CARF-related effectors to sense 3'-5'- and 2'-5'-linked cyclic oligonucleotide signals and protect

bacteria from phage infection. *Cell* **182**, 38–49 (2020).

7. 7.

Morehouse, B. R. et al. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).

8. 8.

Cohen, D. et al. Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).

9. 9.

Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat. Rev. Immunol.* **19**, 31–44 (2019).

10. 10.

Schneider, J. & Imler, J. L. Sensing and signalling viral infection in *Drosophila*. *Dev. Comp. Immunol.* **117**, 103985 (2021).

11. 11.

Goto, A. et al. The kinase IKK β regulates a STING- and NF- κ B-dependent antiviral response pathway in *Drosophila*. *Immunity* **49**, 225–234 (2018).

12. 12.

Liu, Y. et al. Inflammation-induced, STING-dependent autophagy restricts Zika virus infection in the *Drosophila* brain. *Cell Host Microbe* **24**, 57–68 (2018).

13. 13.

Hua, X. et al. Stimulator of interferon genes (STING) provides insect antiviral immunity by promoting Dredd caspase-mediated NF- κ B activation. *J. Biol. Chem.* **293**, 11878–11890 (2018).

14. 14.

Donelick, H. M. et al. In vitro studies provide insight into effects of Dicer-2 helicase mutations in *Drosophila melanogaster*. *RNA* **26**, 1847–1861 (2020).

15. 15.

Jin, L. et al. MPYS is required for IFN response factor 3 activation and type I IFN production in the response of cultured phagocytes to bacterial second messengers cyclic-di-AMP and cyclic-di-GMP. *J. Immunol.* **187**, 2595–2601 (2011).

16. 16.

Burdette, D. L. et al. STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478**, 515–518 (2011).

17. 17.

Diner, E. J. et al. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3**, 1355–1361 (2013).

18. 18.

Gao, P. et al. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP–AMP synthase. *Cell* **153**, 1094–1107 (2013).

19. 19.

Kranzusch, P. J., Lee, A. S., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3**, 1362–1368 (2013).

20. 20.

Zhang, X. et al. Cyclic GMP–AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51**, 226–235 (2013).

21. 21.

Cai, H. et al. 2'3'-cGAMP triggers a STING- and NF-κB-dependent broad antiviral response in *Drosophila*. *Sci. Signal.* **13**, eabc4537 (2020).

22. 22.

Wu, X. et al. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* **42**, 8243–8257 (2014).

23. 23.

Martin, M., Hiroyasu, A., Guzman, R. M., Roberts, S. A. & Goodman, A. G. Analysis of *Drosophila* STING reveals an evolutionarily conserved antimicrobial function. *Cell Rep.* **23**, 3537–3550.e6 (2018).

24. 24.

Caygill, E. E. & Brand, A. H. The GAL4 system: a versatile system for the manipulation and analysis of gene expression. *Methods Mol. Biol.* **1478**, 33–52 (2016).

25. 25.

Tanaka, Y. & Chen, Z. J. STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci. Signal.* **5**, ra20 (2012).

26. 26.

Zhong, B. et al. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity* **29**, 538–550 (2008).

27. 27.

Palmer, W. H., Medd, N. C., Beard, P. M. & Obbard, D. J. Isolation of a natural DNA virus of *Drosophila melanogaster*, and characterisation of host resistance and immune responses. *PLoS Pathog.* **14**, e1007050 (2018).

28. 28.

Slavik, K. M. et al. cGAS-like receptors sense RNA and control 3'2'-cGAMP signaling in *Drosophila*. *Nature*, <https://doi.org/10.1038/s41586-021-03743-5> (2021).

29. 29.

Hartmann, R., Justesen, J., Sarkar, S. N., Sen, G. C. & Yee, V. C. Crystal structure of the 2'-specific and double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase. *Mol. Cell* **12**, 1173–1185 (2003).

30. 30.

Gui, X. et al. Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature* **567**, 262–266 (2019).

31. 31.

McFarland, A. P. et al. Sensing of bacterial cyclic dinucleotides by the oxidoreductase RECON promotes NF-κB activation and shapes a proinflammatory antibacterial state. *Immunity* **46**, 433–445 (2017).

32. 32.

Eaglesham, J. B., McCarty, K. L. & Kranzusch, P. J. Structures of diverse poxin cGAMP nucleases reveal a widespread role for cGAS-STING evasion in host-pathogen conflict. *eLife* **9**, e59753 (2020).

33. 33.

Eaglesham, J. B., Pan, Y., Kupper, T. S. & Kranzusch, P. J. Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS–STING signalling. *Nature* **566**, 259–263 (2019).

34. 34.

Hernáez, B. et al. Viral cGAMP nuclease reveals the essential role of DNA sensing in protection against acute lethal virus infection. *Sci. Adv.* **6**, eabb4565 (2020).

35. 35.

Lowey, B. & Kranzusch, P. J. CD-NTases and nucleotide second messenger signaling. *Curr. Biol.* **30**, R1106–R1108 (2020).

36. 36.

Andersen, L. L. et al. Frequently used bioinformatics tools overestimate the damaging effect of allelic variants. *Genes Immun.* **20**, 10–22 (2019).

37. 37.

Andersen, L. L. et al. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *J. Exp. Med.* **212**, 1371–1379 (2015).

38. 38.

Schulz, A., Jankowski, V., Zidek, W. & Jankowski, J. Highly sensitive, selective and rapid LC-MS method for simultaneous quantification of diadenosine polyphosphates in human plasma. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **961**, 91–96 (2014).

39. 39.

Wang, Y., Holleufer, A., Gad, H. H. & Hartmann, R. Length dependent activation of OAS proteins by dsRNA. *Cytokine* **126**, 154867 (2020).

40. 40.

Holleufer, A. & Hartmann, R. A highly sensitive anion exchange chromatography method for measuring cGAS activity in vitro. *Bio Protoc.* **8**, e3055 (2018).

41. 41.

Dalskov, L. et al. SARS-CoV-2 evades immune detection in alveolar macrophages. *EMBO Rep.* **21**, e51252 (2020).

Acknowledgements

We thank K. Slavik and P. Kranzusch for discussing and sharing information before publication; P. E. Andersen for help with generating Sting-knockout S2 cells; D. Obbard for providing Kallithea virus; and C. Meignin, J. Marques, J. Schneider and G. Haas for helpful discussions. R.H was supported by grants from the Novo Nordisk Foundation (NNF17OC0028184) and the Danish Council for Independent Research (4183-0032B and 0135-00338B). J.-L.I. was supported by Agence Nationale de la Recherche (ANR-17-CE15-0014), Investissement d’Avenir Programs (ANR-10-LABX-0036, ANR-11-EQPX-0022), Institut Universitaire de France and the Chinese National Overseas Expertise Introduction Center for Discipline Innovation (Project ‘111’ (D18010)). H.C was supported by the Natural Science Foundation (32000662) and the Foreign Experts Program (2020A1414010306). A.P. was supported by an ERC consolidator grant (ERC-CoG ProDAP, 817798) and grants from the German Research Foundation (PI 1084/5, TRR179 and TRR237).

Author information

Author notes

1. These authors contributed equally: Andreas Holleufer, Kasper Grønbjerg Winther, Hans Henrik Gad

Affiliations

1. Department of Molecular Biology and Genetics, Aarhus University,
Aarhus, Denmark

Andreas Hollefer, Hans Henrik Gad, Ninna Ahlmann
Frederiksen, Bine Simonsen, Louise Dalskov & Rune Hartmann

2. University of Strasbourg, CNRS UPR9022, Strasbourg, France

Kasper Grønbjerg Winther & Jean-Luc Imler

3. Sino-French Hoffmann Institute, School of Basic Medical Science,
Guangzhou Medical University, Guangzhou, China

Xianlong Ai, Yuqiang Chen, Lihua Li, Ziming Wei, Huimin
Deng, Jiyong Liu, Hua Cai & Jean-Luc Imler

4. Institute of Virology, Technical University of Munich, Munich,
Germany

Line Lykke Andersen & Andreas Pichlmair

5. Bavarian Center for Biomolecular Mass Spectrometry, Technical
University of Munich, Freising, Germany

Karin Kleigrewe

6. German Center for Infection Research (DZIF), Munich partner site,
Munich, Germany

Andreas Pichlmair

Contributions

A.H., K.G.W., H.H.G., H.C., J.-L.I. and R.H. conceived and designed experiments for this study. A.H., K.G.W., H.H.G., X.A., Y.C., L.L., Z.W., H.D., J.L., N.A.F., B.S., L.L.A., K.K., L.D. and H.C. performed experiments. A.H., H.H.G. and H.C. created the figures. H.C., A.P., J.-L.I. and R.H. supervised the study. A.H., J.-L.I. and R.H. wrote the original

draft of the manuscript and all authors participated in reviewing and editing it.

Corresponding authors

Correspondence to [Hua Cai](#) or [Jean-Luc Imler](#) or [Rune Hartmann](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Osamu Nureki 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 Alignment of mammalian cGAS and OAS1 proteins and cGAS-like proteins from *D. melanogaster*.](#)

a, Extended view of the alignment shown in Fig. [1a](#) with the positions of the GS/GG duplet, the metal ion coordinating acidic residues and the Zn finger indicated above the sequences with stars, arrows and bar, respectively. The annotated start methionine in the NCBI reference sequence for isoform D of CG30424 (cGLR2), which would delete the entire spine helix as well as part of the active site, is highlighted in yellow. Deletion of the spine helix is incompatible with a folded enzyme. Furthermore, the GS-containing loop, which coordinates the γ -phosphate of

the donor nucleotide, is a universally conserved feature of nucleotidyltransferases. **b**, **c**, Structure (Protein Data Bank code 4K98) of mouse cGAS in complex with DNA and a cGAMP intermediate analogue (represented in sticks) and two magnesium ions (yellow spheres). The side chains of the acidic active site residues and the serine in the GS motif are also represented in sticks and their oxygen atoms are coloured red. Grey colouring represents the proportion of the protein upstream of the annotated start methionine site in CG30424 (cGLR2). **b**, Full view of the structure. **c**, Enhanced view of the active site.

[Extended Data Fig. 2 Transient expression of candidate cGLRs in S2 cells.](#)

a, Owing to issues with the anti-Flag M2 antibody giving rise to nonspecific bands when performing immunoblots on S2 cell lysates, we replaced the Flag tag we initially used with a V5 tag and reproduced the experiment from Fig. 1b. Cells were transfected with cGAS, cGLR1, cGLR2 and CG7194 or mutants thereof, as well as plasmids encoding firefly or *Renilla* luciferase under transcriptional control of the *Sting* or *Actin5C* promoter, respectively. At 24 h after transfection, luciferase activity was measured. Data are from one experiment performed in biological triplicates and are shown with mean \pm s.d. ($n = 3$). **b**, Immunoblot showing the expression of the V5-tagged proteins from **a**. cGLR2 appears to be rapidly degraded in S2 cells, whereas it could easily be detected in HEK293T cells following transfection (Extended Data Fig. 10a). For gel source data, see Supplementary Fig. 1.

[Extended Data Fig. 3 Ectopic expression of cGLR1 or cGLR2 in transgenic flies induces expression of *Srg2* and *Srg3*.](#)

a–d, Expression of *cGLR1* (**a**), *cGLR2* (**b**), *Srg2* (**c**) and *Srg3* (**d**) was monitored by RT–qPCR in transgenic flies ectopically expressing wild-type or mutant cGLRs. Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments (red, blue and green, each performed in biological triplicate) and shown with mean ($n = 9$). **e**, Flies of the indicated genotypes were injected with Tris and survival was

monitored daily. Data are from 3 independent experiments, each with 30 flies ($n = 90$). In **a–d**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test and compared to relevant AFA mutants. In **e**, log-rank test was used to compare cGLR1 and cGLR1 AFA ($P = 0.0030$) and cGLR2 and cGLR2 AFA ($P = 0.0109$)

[Source data](#).

[Extended Data Fig. 4 cGLR signalling depends on Sting and Relish.](#)

a, Sanger sequencing showing the 2-bp deletion in the *Sting* gene in Sting-knockout S2 cells. **b**, Immunoblot showing the lack of expression of Sting in Sting-knockout S2 cells. The arrow indicates the position of the Sting band. The presented immunoblot is representative out of two independent immunoblots. For gel source data, see Supplementary Fig. 1. **c, d**, S2 cells transfected with expression vectors for cGAS, cGLR1 or cGLR2 and plasmids encoding firefly or *Renilla* luciferase under transcriptional control of the *Sting* or *Actin5C* promoter, respectively, were used to monitor activation of the Sting pathway. **c**, S2 cells with or without co-transfection with an expression vector for Sting. **d**, Sting-knockout S2 cells with or without co-transfection with an expression vector for Sting. **c, d**, Data are from three independent experiments (blue, red, and green, each performed in biological triplicates) and are shown with mean ($n = 9$). Data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test. **e**, S2 cells transfected with expression vectors for cGAS, cGLR1 or cGLR2 and plasmid encoding firefly luciferase under transcriptional control of the *Sting* promoter or a mutated version containing two point mutations in the Relish binding site. Data are from three independent experiments (blue, red and green, each performed in biological triplicates) and are shown with mean ($n = 9$). Data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test, in which the groups were pairwise compared to mock.

[Extended Data Fig. 5 Generation of cGLR1- and cGLR2-knockout flies.](#)

The *cGLR1* and *cGLR2* genes, both located on the right arm of the second chromosome, are shown together with their annotated transcripts. Open reading frames are indicated in green. For *cGLR1*, an 8-bp deletion was introduced in the first exon using CRISPR–Cas9 technology. The deletion creates a frameshift after the asparagine residue at position 31, leading to termination of translation after insertion of single glycine residue. For *cGLR2*, a 5-bp deletion was created in exon 3, which is shared by all isoforms. The deletion results in a frameshift after the glutamate residue at position 338, leading to termination of translation after insertion of a 32 amino acid extension (HDRRIDPGSSLGNVPVRAKDSKRPEGRRDQPE).

[Extended Data Fig. 6 Expression of *Sting*, *Srg2* and *Srg3* in flies upon infection with DCV.](#)

a, Corresponding control to Fig. 2a. *w¹¹¹⁸*, cGLR1-knockout, cGLR2-knockout or cGLR1/2-knockout flies were injected with Tris and survival was monitored daily. Data are from 3 independent experiments, each with 3 groups of around 10 flies, and shown with s.e.m. **b–d**, *w¹¹¹⁸*, cGLR1-knockout, cGLR2-knockout or cGLR1/2-knockout flies were injected with DCV or Tris and expression of *Sting* (**b**), *Srg2* (**c**) and *Srg3* (**d**) was monitored by RT–qPCR at 2 and 3 days after injection (days post-infection (dpi)). Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments, each performed in biological triplicates ($n = 9$) shown with mean \pm s.e.m. In **a**, log-rank test was used to test whether the survival curves differed ($P = 0.0569$). In **b–d**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test

[Source data](#).

[Extended Data Fig. 7 Expression of *Sting*, *Srg2* and *Srg3* in flies upon infection with Kallithea virus.](#)

a, Corresponding control to Fig. 2b. *w¹¹¹⁸*, cGLR1-knockout or cGLR2-knockout flies were injected with Tris and survival was monitored daily.

Data are from 3 independent experiments, each with 3 groups of around 10 flies, and shown with s.e.m. **b–d**, w^{1118} , cGLR1-knockout or cGLR2-knockout flies were injected with Kallithea virus or Tris and expression of *Sting* (**b**), *Srg2* (**c**) and *Srg3* (**d**) was monitored by RT-qPCR at 5 and 10 days after injection. Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments, each performed in biological triplicates ($n = 9$) shown with mean \pm s.e.m. **e, f**, w^{1118} , cGLR1-knockout, cGLR2-knockout or cGLR1/2-knockout flies were injected with Kallithea virus (**e**) or Tris (**f**) and survival was monitored daily. Data are from 3 independent experiments, each with 3 groups of around 10 flies, and shown with s.e.m. **g**, w^{1118} , cGLR1-knockout, cGLR2 or cGLR1/2-knockout flies were injected with Kallithea virus and viral load was monitored by RT-qPCR at 5 days after infection. Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments, each performed in triplicates ($n = 9$) shown with mean \pm s.e.m. In **a, e, f**, log-rank test was used to compare the survival curves pairwise followed by a Holm–Šídák multiple comparison correction (**a, e**) or all and once (**f**) ($P = 0.1452$). In **b–d**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test. In **g**, log-transformed data were analysed using one-way ANOVA and a two-tailed Dunnett T3 post hoc test and compared to w^{1118} flies

[Source data.](#)

Extended Data Fig. 8 Loss of cGLR1 or cGLR2 has a limited effect on VSV infection in flies.

a, b, w^{1118} , cGLR1-knockout or cGLR2-knockout flies were injected with VSV (**a**) or Tris (**b**) and survival was monitored daily. Data are from 3 independent experiments, each with 3 groups of around 10 flies, and shown with s.e.m. **c–g**, w^{1118} , cGLR1-knockout, cGLR2-knockout or cGLR1/2-knockout flies were injected with VSV or Tris and viral load (**c**) as well as expression of *Sting* (**d**), *Srg1* (**e**), *Srg2* (**f**) and *Srg3* (**g**) were monitored by RT-qPCR at 4 and 5 days after infection. Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments, each performed in biological triplicates ($n = 9$) shown with mean \pm s.e.m. In

a, b, log-rank test was used to test whether the survival curves differed (**a**, $P = 0.0772$; **b**, $P = 0.1942$). In **c**, log-transformed data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test and compared to w^{1118} flies. In **d–g**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test

[Source data](#).

Extended Data Fig. 9 Loss of cGLR1 or cGLR2 has a limited effect on IIV6 infection in flies.

a, b, w^{1118} , cGLR1-knockout or cGLR2-knockout flies were injected with IIV6 (**a**) or Tris (**b**) and survival was monitored daily. Data are from 3 independent experiments, each with 3 groups of around 10 flies, and shown with s.e.m. **c–g**, w^{1118} , cGLR1-knockout or cGLR2-knockout flies were injected with IIV6 or Tris and viral load (**c**) as well as expression of *Sting* (**d**), *Srg1* (**e**), *Srg2* (**f**) and *Srg3* (**g**) were monitored by RT–qPCR at 5 and 10 days after infection. Expression was normalized to the housekeeping gene *RpL32*. Data are from three independent experiments, each performed in biological triplicates ($n = 9$) shown with mean \pm s.e.m. In **a, b**, log-rank test was used to test whether the survival curves differed (**a**, $P = 0.3536$; **b**, $P = 0.7337$). In **c**, log-transformed data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test and compared to w^{1118} flies. In **d–g**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test

[Source data](#).

Extended Data Fig. 10 cGLR2 produce 2'3'-cGAMP and 3'2'-cGAMP, which can activate human and *Drosophila* STING.

a, Immunoblot showing the expression of STING and Flag-tagged nucleotidyltransferases from Fig. [3a](#). **b**, HEK293T cells transfected with cGAS, cGAS AFA, cGLR2, cGLR2 AFA and STING as indicated and plasmids encoding firefly or *Renilla* luciferase under transcriptional control of the *IFNB1* or a constitutive promoter, respectively. Data are from three

independent experiments (blue, red and green, each performed in biological triplicates) and shown with mean ($n = 9$). **c**, Immunoblot showing the expression of STING and Flag-tagged nucleotidyltransferases from **b**. For gel source data, see Supplementary Fig. 1. **d**, Representative chromatograms from mass spectrometry analysis of 2'3'-cGAMP or 3'2'-cGAMP spiked lysates from GFP-transfected cells. **e**, Representative chromatogram from mass spectrometry analysis of cGAS-transfected cells. **f**, Representative chromatograms from mass spectrometry analysis of GFP- or cGLR2-transfected cells. **g–i**, HT-1080 (**g**) or S2 cells (**h**, **i**) permeabilized using digitonin and treated with the indicated cGAMPs. Expression of *IFNB1* (**g**), *Sting* (**h**) or *Srg3* (**i**) were monitored by RT-qPCR at 8 and 6 h after treatment for HT-1080 and S2 cells, respectively. Expression was normalized to the housekeeping genes *GADPH* (**g**) and *RpL32* (**h**, **i**). Data are shown with mean \pm s.e.m. ($n = 4$ for **g**; $n = 3$ for **h**, **i**) and are biological replicates from one representative experiment out of two independent experiments. In **b**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test and compared to mock. In **g–i**, data were analysed using one-way ANOVA and a two-tailed Dunnett's T3 post hoc test and compared to mock.

[Extended Data Fig. 11 cGLR1 produces 3'2'-cGAMP in response to dsRNA.](#)

a, Representative chromatograms from anion exchange chromatography analysis of different nucleotides. **b**, Representative chromatograms from anion exchange chromatography analysis of reaction products from activity assays with recombinant cGLR1 in the presence of different nucleic acids. **c**, Immunoblot showing the expression of Flag-tagged cGLR2 and mutants thereof from Fig. 3f. For gel source data, see Supplementary Fig. 1. **d**, S2 cells transfected with cGLR2 or mutants thereof as well as plasmids encoding firefly or *Renilla* luciferase under transcriptional control of the *Sting* or *Actin5C* promoter, respectively. Data are from three independent experiments (blue, red and green, each performed in biological triplicates) and shown with mean ($n = 9$). In **d**, data were analysed using two-way ANOVA and a two-tailed Holm–Šídák post hoc test.

Supplementary information

Supplementary Figure

This file contains the uncropped gel source data for Extended Data Figs 2, 4, 10 and 11.

Reporting Summary

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Holleufer, A., Winther, K.G., Gad, H.H. *et al.* Two cGAS-like receptors induce antiviral immunity in *Drosophila*. *Nature* **597**, 114–118 (2021). <https://doi.org/10.1038/s41586-021-03800-z>

- Received: 19 February 2021
- Accepted: 05 July 2021
- Published: 14 July 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03800-z>

cGAS-like receptors put a sting into the evolution of immune defence

- Cara West
- Neal Silverman

News & Views 09 Aug 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03800-z>

- Article
- [Published: 25 August 2021](#)

A clinically applicable integrative molecular classification of meningiomas

- [Farshad Nassiri^{1,2,3}](#),
- [Jeff Liu^{1,3}](#) na1,
- [Vikas Patil^{1,3}](#) na1,
- [Yasin Mamatjan](#) [ORCID: orcid.org/0000-0002-8807-2927^{1,3}](#) na1,
- [Justin Z. Wang^{1,2,3}](#) na1,
- [Rupert Hugh-White⁴](#),
- [Andrew M. Macklin⁵](#),
- [Shahbaz Khan⁵](#),
- [Olivia Singh¹](#),
- [Shirin Karimi¹](#),
- [Rosario I. Corona](#) [ORCID: orcid.org/0000-0003-0470-4446⁴](#),
- [Lydia Y. Liu](#) [ORCID: orcid.org/0000-0001-6026-3169^{4,14}](#),
- [Caroline Y. Chen](#) [ORCID: orcid.org/0000-0002-7552-6483⁴](#),
- [Ankur Chakravarthy⁵](#),
- [Qingxia Wei¹](#),
- [Bharati Mehani](#) [ORCID: orcid.org/0000-0002-6198-7692⁶](#),
- [Suganth Suppiah^{1,2,3}](#),
- [Andrew Gao⁷](#),
- [Adriana M. Warkewych¹](#),
- [Ghazaleh Tabatabai^{3,8}](#),
- [Paul C. Boutros](#) [ORCID: orcid.org/0000-0003-0553-7520^{3,4,9}](#),
- [Gary D. Bader](#) [ORCID: orcid.org/0000-0003-0185-8861^{10,11,12,13}](#),
- [Daniel D. de Carvalho^{5,14}](#),

- [Thomas Kislinger](#) ORCID: [orcid.org/0000-0003-3525-5540^{3,5,14}](https://orcid.org/0000-0003-3525-5540),
- [Kenneth Aldape](#)^{1,3,6 na2} &
- [Gelareh Zadeh](#) ORCID: [orcid.org/0000-0002-6637-4502^{1,2,3,5 na2}](https://orcid.org/0000-0002-6637-4502)

[Nature](#) volume **597**, pages 119–125 (2021)

- 4006 Accesses
- 111 Altmetric
- [Metrics details](#)

Subjects

- [Cancer epigenetics](#)
- [Cancer genomics](#)
- [CNS cancer](#)

Abstract

Meningiomas are the most common primary intracranial tumour in adults¹. Patients with symptoms are generally treated with surgery as there are no effective medical therapies. The World Health Organization histopathological grade of the tumour and the extent of resection at surgery (Simpson grade) are associated with the recurrence of disease; however, they do not accurately reflect the clinical behaviour of all meningiomas². Molecular classifications of meningioma that reliably reflect tumour behaviour and inform on therapies are required. Here we introduce four consensus molecular groups of meningioma by combining DNA somatic copy-number aberrations, DNA somatic point mutations, DNA methylation and messenger RNA abundance in a unified analysis. These molecular groups more accurately predicted clinical outcomes compared with existing classification schemes. Each molecular group showed distinctive and prototypical biology (immunogenic, benign *NF2* wild-type, hypermetabolic and proliferative) that informed therapeutic options. Proteogenomic characterization reinforced the robustness of the newly defined molecular groups and uncovered highly abundant and group-specific protein targets

that we validated using immunohistochemistry. Single-cell RNA sequencing revealed inter-individual variations in meningioma as well as variations in intrinsic expression programs in neoplastic cells that mirrored the biology of the molecular groups identified.

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: Integrative multiplatform analysis reveals four molecular groups of meningioma.

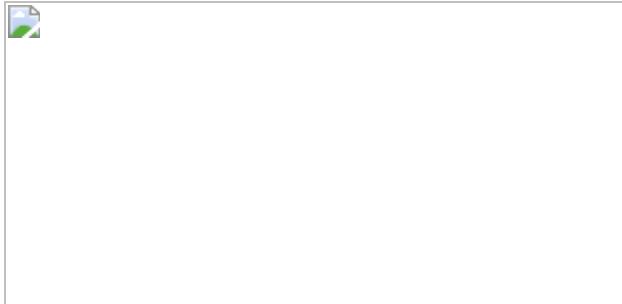


Fig. 2: Molecular groups are distinguished by prototypical biology that inform on new therapeutics.

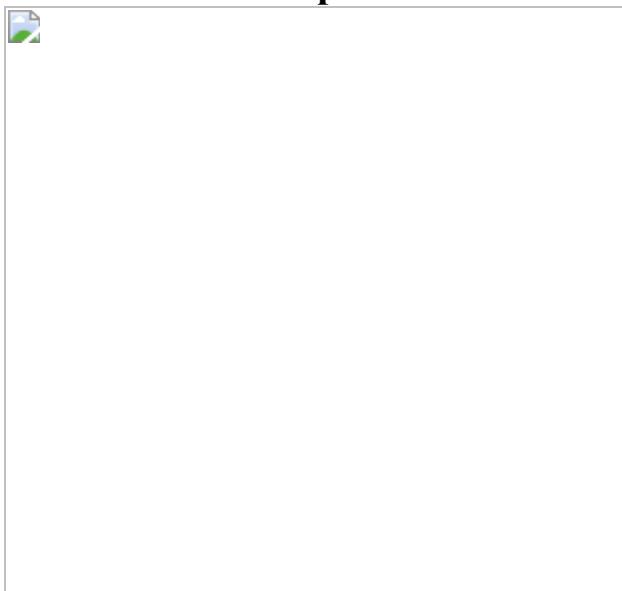


Fig. 3: Proteogenomic characterization validates the robustness of molecular groups and identifies markers that can distinguish molecular groups by immunohistochemistry.

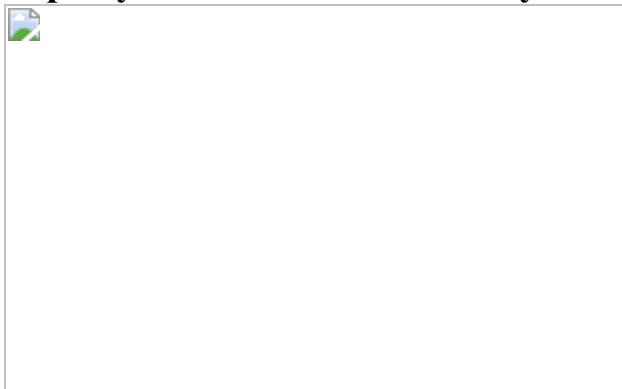


Fig. 4: Single-cell RNA sequencing of human meningiomas reveals substantial inter-patient heterogeneity and subtle within-patient

variability.



Data availability

Raw sequencing data for all datatypes have been deposited into public repositories. Proteomic data has been deposited to the Mass Spectrometry Interactive Virtual Environment (MassIVE, <https://massive.ucsd.edu/>; ID MSV000086901). DNA methylation idat files have been deposited to the Gene Expression Omnibus (GEO; [GSE180061](https://www.ncbi.nlm.nih.gov/geo/study/GSE180061)). Whole-exome sequencing (fastq), bulk mRNA (fastq) and snRNA (fastq) datasets have been deposited to the European Genome Archive (<https://www.ebi.ac.uk/ega/>) under study ID EGAS00001004982 and dataset IDs EGAD00001007051, EGAD00001007494 and EGAS00001004982. The processed genomic data has been submitted to cBioportal at https://www.cbioportal.org/study/summary?id=mng_utoronto_2021. [Source data](#) are provided with this paper.

Code availability

Specific code will be made available upon request to G.Z.

References

1. 1.

Ostrom, Q. T. et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro. Oncol.* **21**, v1–v100 (2019).

2. 2.

Goldbrunner, R. et al. EANO guidelines for the diagnosis and treatment of meningiomas. *Lancet Oncol.* **17**, e383–e391 (2016).

3. 3.

Sahm, F. et al. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol.* **18**, 682–694 (2017).

4. 4.

Clark, V. E. et al. Genomic analysis of non-*NF2* meningiomas reveals mutations in *TRAF7*, *KLF4*, *AKT1*, and *SMO*. *Science* **339**, 1077–1080 (2013).

5. 5.

Brastianos, P. K. et al. Genomic sequencing of meningiomas identifies oncogenic *SMO* and *AKT1* mutations. *Nat. Genet.* **45**, 285–289 (2013).

6. 6.

Clark, V. E. et al. Recurrent somatic mutations in *POLR2A* define a distinct subset of meningiomas. *Nat. Genet.* **48**, 1253–1259 (2016).

7. 7.

Harmancı, A. S. et al. Integrated genomic analyses of de novo pathways underlying atypical meningiomas. *Nat. Commun.* **8**, 14433 (2017).

8. 8.

Nassiri, F. et al. DNA methylation profiling to predict recurrence risk in meningioma: development and validation of a nomogram to optimize clinical management. *Neuro. Oncol.* **21**, 901–910 (2019).

9. 9.

Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018).

10. 10.

Koboldt, D. C. et al. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).

11. 11.

Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).

12. 12.

Hoadley, K. A. et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).

13. 13.

Reshef, D. N. et al. Detecting novel associations in large data sets. *Science* **334**, 1518–1524 (2011).

14. 14.

Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).

15. 15.

Teschendorff, A. E. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. *Genome Med.* **12**, 56 (2020).

16. 16.

Yang, Z. et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* **17**, 205 (2016).

17. 17.

Silva, T. C. et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* **35**, 1974–1977 (2019).

18. 18.

Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

19. 19.

Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

20. 20.

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

21. 21.

McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

22. 22.

Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).

23. 23.

Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

24. 24.

Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

25. 25.

Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).

26. 26.

McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

27. 27.

Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* 1–16 (2017).

28. 28.

Chang, M. T. et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34**, 155–163 (2016).

29. 29.

Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).

30. 30.

Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).

31. 31.

Cerami, E. et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

32. 32.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

33. 33.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

34. 34.

Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

35. 35.

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).

36. 36.

Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).

37. 37.

Dentro, S. C., Wedge, D. C. & Van Loo, P. Principles of reconstructing the subclonal architecture of cancers. *Cold Spring Harb. Perspect. Med.* **7**, a026625 (2017).

38. 38.

Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

39. 39.

Reimand, J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protocols* **14**, 482–517 (2019).

40. 40.

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).

41. 41.

Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2013).

42. 42.

Kent, W. J. BLAT—the BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).

43. 43.

Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).

44. 44.

Wingender, E., Schoeps, T., Haubrock, M., Krull, M. & Dönitz, J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* **46**, D343–D347 (2018).

45. 45.

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).

46. 46.

The 1000 Genomes Project Consortium A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

47. 47.

Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).

48. 48.

L. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

49. 49.

Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122 (2016).

50. 50.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

51. 51.

Tickle, T. I., Georgescu, C., Brown, M. & Haas, B. Infer copy number variation from single-cell RNA-seq data.
<https://doi.org/10.1101/29.bioc.infercnv> (2019).

52. 52.

Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

53. 53.

Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018).

54. 54.

Jerby-Arnon, L. et al. Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).

55. 55.

Kinker, G. S. et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).

56. 56.

Izar, B. et al. A single-cell landscape of high-grade serous ovarian cancer. *Nat. Med.* **26**, 1271–1279 (2020).

57. 57.

Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

58. 58.

Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).

Acknowledgements

F.N. is supported by the Canadian Institute of Health Research (CIHR) Vanier Scholarship, AANS/CNS Section on Tumors & NREF Research Fellowship Grant, and Hold’em for Life Oncology Fellowship. G.Z is funded by a CIHR Project Grant Award (173241), a Canadian Cancer Society Innovation Grant (706898) as well as a Quest for Cures Grant (GN-000430) and Clinical Biomarkers Grant (GN-000693) from the Brain Tumour Charity UK. P.C.B. was supported by the NIH/NCI under awards P30CA016042, U24CA248265 and P50CA211015. We also thank N. Dewitt for editing, and the Mary Hunter Meningioma Program and Toronto Western Hospital Foundation.

Author information

Author notes

1. These authors contributed equally: Jeff Liu, Vikas Patil, Yasin Mamatjan, Justin Z. Wang
2. These authors jointly supervised this work: Kenneth Aldape, Gelareh Zadeh

Affiliations

1. MacFeeters Hamilton Neuro-Oncology Program, Princess Margaret Cancer Centre, University Health Network and University of Toronto,

Toronto, Ontario, Canada

Farshad Nassiri, Jeff Liu, Vikas Patil, Yasin Mamatjan, Justin Z. Wang, Olivia Singh, Shirin Karimi, Qingxia Wei, Suganth Suppiah, Adriana M. Workewych, Kenneth Aldape & Gelareh Zadeh

2. Division of Neurosurgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada

Farshad Nassiri, Justin Z. Wang, Suganth Suppiah & Gelareh Zadeh

3. The International Consortium on Meningiomas, Toronto, Ontario, Canada

Farshad Nassiri, Jeff Liu, Vikas Patil, Yasin Mamatjan, Justin Z. Wang, Suganth Suppiah, Ghazaleh Tabatabai, Paul C. Boutros, Thomas Kislinger, Kenneth Aldape & Gelareh Zadeh

4. Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

Rupert Hugh-White, Rosario I. Corona, Lydia Y. Liu, Caroline Y. Chen & Paul C. Boutros

5. Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

Andrew M. Macklin, Shahbaz Khan, Ankur Chakravarthy, Daniel D. de Carvalho, Thomas Kislinger & Gelareh Zadeh

6. Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA

Bharati Mehani & Kenneth Aldape

7. Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada

Andrew Gao

8. Department of Neurology and Interdisciplinary Neuro-Oncology, Hertie Institute for Clinical Brain Research, Center for Neuro-Oncology, Comprehensive Cancer Center, University Hospital Tübingen, Tübingen, Germany

Ghazaleh Tabatabai

9. Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

Paul C. Boutros

10. The Donnelly Center, University of Toronto, Toronto, Ontario, Canada

Gary D. Bader

11. Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Gary D. Bader

12. Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

Gary D. Bader

13. The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada

Gary D. Bader

14. Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

Lydia Y. Liu, Daniel D. de Carvalho & Thomas Kislinger

Contributions

F.N., K.A. and G.Z. conceived and designed the study. F.N., S.S., and J.Z.W. collected all biomaterials and clinical data. K.A., A.G. and S. Karimi reviewed the pathological sections. F.N. prepared specimens for whole-exome sequencing, DNA methylation, mRNA sequencing and single-cell RNA sequencing. S. Khan and A.M. carried out proteomic experiments. O.S., S. Karimi and F.N. carried out immunohistochemical experiments and analyses. F.N., J.Z.W. and Q.W. carried out in vitro and in vivo experimentation. F.N., J.L., Y.M., V.P., A.C., R.H.-W., R.I.C., L.Y.L., C.Y.C. and B.M. contributed to the data processing and analyses. F.N., K.A., P.C.B., G.D.B., D.D.d.C., T.K. and G.Z. contributed to data interpretation. F.N. and A.M.W. organized the figures. F.N. and G.Z. wrote the first draft as well as subsequent revisions and the response to reviewers. All authors contributed to the final data interpretation and critical revision of the manuscript and approved the final version of the manuscript. G.Z. supervised all aspects of the study.

Corresponding author

Correspondence to [Gelareh Zadeh](#).

Ethics declarations

Competing interests

D.D.d.C. and A.C. are listed as inventors on patents filed that are unrelated to this project. D.D.d.C. received research funding from Pfizer and Nektar therapeutics that was not related to this project. D.D.d.C is co-founder, shareholder and CSO of Adela, Inc. P.C.B sits on the Scientific Advisory Boards of BioSymetrics Inc. and Intersect Diagnostics Inc. G.T. has served on advisory boards of AbbVie, Bayer and BMS; received consulting fees from AbbVie, Bayer; received speaker fees from Medac and Novocure; received travel grants from Novocure, Medac and BMS; received research grants from Roche Diagnostics and Medac, all not related to this work.

Additional information

Peer review information *Nature* thanks Itay Tirosh, Roel Verhaak 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 Individual datatype classification of meningiomas.](#)

a, Violin plots showing the distribution of the normalized mutual information (MI) for each pairwise comparison of datatype. Median is shown as white dot. The number of total genes and number of genes with statistically significant ($FDR < 5\%$) MI values are shown. Below this is a heatmap showing the consensus clustering of genes where MI was significant for at least one datatype pair. Rows represent a gene for which data exists from all data types. **b,d,f**, Unsupervised consensus hierarchical clustering of (b), 5,000 genes that show that highest median absolute deviation across expression values, (d), 10,000 CpG sites that show that highest median absolute deviation across β -values, (f), 1,000 genes that show that highest median absolute deviation across copy number ratios. Heatmap of consensus matrices with $K = 6$ groups (b,d,f) are displayed. Overall, six groups were most stable across all platforms. **c,e,g**, Kaplan Meier-plot displaying recurrence-free survival (RFS) distributions of unsupervised cluster assignments by (c) mRNA data, (e) DNA methylation data, (g) copy number data. The associations with outcomes are unique for the 6 cluster groups obtained on individual platform analyses. **h**, Average silhouette widths for unsupervised consensus hierarchical clustering from $K = 2$ to $K = 10$. The silhouette score is a measure of stability of number of groups. Higher scores indicate greater stability and robustness. Average silhouette width is highest at $K = 4$ subgroups. **i**, Alluvial plot demonstrating associations between WHO grade and integrative molecular groups defined in this study. **j-l**, Kaplan Meier-plot displaying recurrence-free survival (RFS) distributions of patients stratified and colored by

molecular group assignments for WHO grade 1 tumors (j), WHO grade 2 tumors (k), and WHO grade 3 tumors (l).

Extended Data Fig. 2 Generalizability of the association of molecular groups with outcome.

a, Ensemble of Receiver Operating Characteristic (ROC) curves from 50 iterations of trained MG-versus-other models. Overlaid for each model is the mean Area Under the Curve (AUC) and its associated 95% confidence interval for samples in corresponding test sets. **b**, Heatmap showing results of single-sample Gene-Set Enrichment Analysis (ssGSEA) using mRNA data in an independent cohort of 80 meningioma samples. Each sample in the validation set was assigned a score for molecular groups 1, 2, 3 and 4 using gene-expression based signatures from the discovery cohort. MG designation was determined by highest scores from ssGSEA assignments. Unsupervised hierarchical clustering using scores from MG assignments revealed four distinctive groups of tumors with 97% of samples having concordant assignment by maximal scores. Samples almost always showed high scores that were distinctive to only a single group, highlighting the robustness of classification in an independent cohort. **c**, Brier prediction curve for recurrence-free survival comparing molecular group to WHO grade in the generalization cohort. The models tested were those developed on the discovery cohort. Prediction errors are consistently lowest using molecular groups in comparison to the validation cohort. **d**, Kaplan Meier-plot displaying recurrence-free survival (RFS) distribution of patients stratified and colored by molecular group assignments for generalization set. P value reported is a Log Rank Test. Distributions are highly similar to those obtained in discovery cohort.

Extended Data Fig. 3 Most mutations are clonal in meningioma.

a, Lollipop plots showing the distribution of NF2 mutations by genomic regions within each molecular group. **b**, Mutational burden (nonsynonymous mutations per megabase) of meningiomas stratified by molecular groups in comparison to other TCGA solid cancers. Every dot

represents a sample and horizontal lines are median number of mutations in each cancer type. Mutational burden in each cancer is ordered by percentile rank. Cancer types are ordered on the horizontal axis based on their median numbers of somatic mutations. Mutational burden of MG4 tumors is statistically higher than tumours in MG1-3 ($P = 1.6 \times 10^{-3}$, Kruskal Wallis test). **c**, Distribution of the number of mutations that are considered clonal per each patient sample (column). A total of 26% of tumors exhibited only clonal point mutations. In the median tumor, 75% of single nucleotide variants were clonal. **d**, Cancer cell fraction of all variants in each patient sample (columns) ordered as in (c). Variants are colored according to the classification in the legend. **e**, Cancer cell fraction of recurrent oncogenic driver mutations (columns). Variants are colored according to the classification in the legend.

[Extended Data Fig. 4 Genomic disruptions differ among molecular groups.](#)

a, Genome-wide copy-number alterations computed from whole-exome sequencing data. Arrangements of copy number profile are matched to the samples from mutation plot above. Only mutations that are relevant to discussion in text are shown. **b**, Boxplots showing the mRNA expression of NF2 stratified by molecular group. Each dot is a sample. Samples are colored by NF2 mutation status and shapes are according to NF2 deletion status by CNA. Some MG3 and MG4 meningiomas that are NF2 wildtype show silencing of NF2 expression. **c**, Boxplots comparing the mean methylation level of NF2 wildtype MG3 and MG4 meningiomas with high versus low NF2 expression using all probes (left), those mapping to the promoter region (middle), and those mapping to the gene body (right). **d**, Circos plot showing the landscape of interchromosomal gene rearrangements detected using a stringent threshold for conservative estimation of fusion events (unique spanning reads ≥ 25) in each molecular group. Total number of interchromosomal fusion in MG1, MG2, MG3 and MG4 are 2, 7, 18, and 23, respectively.

[Extended Data Fig. 5 Gene expression profiles of molecular groups.](#)

a, Hierarchical clustering of the expression of genes from select pathways identified in Fig. 2a. Selected genes have been labeled. Redundancy of genes to pathways is shown in the side bar. **b**, Boxplots showing the results for estimates of immune and stromal infiltration by DNA methylation (LUMP score on left and methylCIBERSORT in middle) and somatic DNA alterations (right, ABSOLUTE score). **c**, Scatterplots comparing normalized enrichment scores between molecular groups using Gene Set Variation Analysis (GSVA). Each dot is a pathway. Shown at the top of each panel are Pearson correlations and associated 95%CI. MG2 tumors were divided into tumors that are driven by CNA (MG2-CNA) and tumors that are driven by mutations (MG2-Mut). Correlations were highest when comparing MG2 tumors driven by CNA to MG2 tumors driven by mutations (red box). **d**, Hierarchical clustering of normalized enrichment scores from (c) identifies MG2-CNA and MG2-Mut tumors as one coherent group. **e**, Boxplots comparing the activation of molecular signatures of proliferation between MGs. Statistical significance is denoted by asterisks.

Extended Data Fig. 6 Molecular characterization of patient derived cell lines.

a, t-distributed Stochastic Neighbor Embedding (tSNE) plot of genome-wide DNA methylation profiles of patient derived cell lines (red), to meningiomas (blue), and 2798 previously published tumors from 40 other brain tumor types⁵⁸. **b**, Heatmap showing results of single-sample Gene-Set Enrichment Analysis (ssGSEA) using mRNA data from cell lines. Each cell line was assigned a score for molecular groups 1, 2, 3 and 4 using gene-expression based signatures from the discovery cohort. Molecular group designation was determined by highest scores from ssGSEA assignments. **c**, Gross morphological images of a representative MG4-xenografted mice. Extra axial tumor is outlined in dashed yellow lines. Compression on adjacent neural structures is evident after partial (middle panel) and complete (right panel) separation of meningioma from brain. **d**, Serial sections and immunostaining for MCM2 in representative MG4-xenografted mice. Scale bar is 2mm. Small areas of tumor that have invaded the brain can be seen staining for MCM2.

Extended Data Fig. 7 Proteomic and gene expression data converge to similar biology driving each molecular group.

a, Hierarchical clustering of normalized enrichment scores obtained by Gene-Set Variation Analysis (GSVA) using proteomic data (rows) and mRNA data (columns). **b**, Distribution of correlation of mRNA expression to protein abundance in all samples (grey), MG1 meningiomas (red), MG2 meningiomas (blue), MG3 meningiomas (green) and MG4 meningiomas (orange). Vertical line indicates overall median correlation across all samples (Spearman's $r = 0.279$, 95%CI 0.273-0.284). **c**, Scatterplots comparing normalized enrichment scores by GSVA using gene expression (x-axis) and protein abundance (y-axis) stratified by MG classifications. Each dot represents a pathway. Pathways that are statistically significant and concordant by protein and mRNA data are colored green while those that are discordant are colored red. Pearson correlations and 95% confidence intervals are indicated at the top of each panel. **d**, Network of activated gene circuits by proteome data in $N = 96$ samples. Protein groups were ranked for each subtype by degree of differential expression. Gene-set enrichment analysis was performed on the ranked gene lists and enriched pathways are visualized using the EnrichmentMap plugin in Cytoscape App. Nodes represent pathways and edges represent shared genes between pathways. Pathways above horizontal line are up-regulated (red nodes) in each molecular group while pathways below horizontal line are down-regulated (blue nodes) in each molecular group.

Extended Data Fig. 8 Differences in genome-wide methylation across meningioma groups.

a, Hierarchical clustering of highly differentially methylated CpGs (absolute $\Delta\beta > 0.35$, FDR < 0.05) between all meningiomas and healthy meninges. Annotations of molecular groups are on the right side of heatmap. **b**, Boxplots showing the distribution of β values for probes in **(a)** that are hypomethylated in healthy meninges (left) and hypermethylated in healthy meninges (right). Pairwise comparisons in each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant). **c**, Boxplots showing the distribution of using epigenetic mitotic

clocks with epiTOC model (left), epiTOC2 model (middle), and HypoClock model (right). Pairwise comparisons in each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant). **d**, Number of unique and overlapping probes that are differentially methylated (absolute $\Delta\beta > 0.1$, FDR < 0.05) when comparing each molecular group to healthy meninges. **e**, Scatterplots comparing master regulator transcription factor expression with average β values at sites enriched for the motif of that transcription factor. Samples are colored according to molecular group. Pearson correlation with 95% confidence intervals are reported. Hypomethylation at motifs of immunological-lineage-specific transcription factors such as PU.1, RUNX1/2 and IRF5/8 were enriched in immunogenic (MG1) tumors ($P = 1.05 \times 10^{-8}$, hypergeometric test) and associated with enhancer hypomethylation. Similarly, master regulators of cell proliferation such as MYBL2, LHX4, and FOXM1 were hypomethylated in proliferative (MG4) tumors and associated with increased abundance of these transcription factors ($P = 1.24 \times 10^{-3}$, hypergeometric test).

Extended Data Fig. 9 Meningiomas show low within patient variation of expression and copy number profile.

a, Pairwise correlations of expression profiles of all cells ordered by hierarchical clustering. Each cell is annotated to tumor of origin from Fig. 4a and cluster assignments from Fig. 4b at top and side bars. **b**, Inferred genome-wide copy number variations of single nuclei of healthy meninges (reference, top panel), immune cells (middle panel), and neoplastic cells (bottom panel). Sample and cluster annotation are shown on the left. The copy number plot of these tumors are homogenous and subclones of cells within tumors with distinct copy number profiles are not common. Annotation to patient of origin and cluster on the left of each heatmap. **c**, Scatterplots showing the relationship between arm-level CNA inferred by snRNA-seq (x-axis) to matched CNA by bulk whole exome sequencing (y-axis). Two representative samples are shown.

Extended Data Fig. 10 The transcriptome of MGs is shaped by the expression profiles from both neoplastic and non-neoplastic

cells.

a, Bubble plot showing the expression of lineage specific markers for distinct cell types. **b**, Stacked barplot showing the relationship of samples to clusters. Samples are colored by patient of origin as in Fig. 4a. Barplot to the right shows the number of cells within each cluster. **c**, The top heatmap shows hierarchical clustering results of single cells by molecular group scores. Each cell was scored for the bulk signature of each molecular group and scores were compared to a permuted random gene set. Shown are cells with at least one score with FDR < 0.2. Scores were scaled such that the sum of all scores for each cell is equal to one. Below is a matched heatmap showing the number of genes detected for each MG signature in each cell. In a subset of cells, low scores are associated with low detection rate of genes (yellow and pink boxes). **d**, Stacked barplot showing the distribution of immune versus non-immune cells across molecular groups (left) and cycling versus non-cycling neoplastic cells across molecular groups (right) to clusters. Samples are colored by molecular group of tumor as in Fig. 4d. **e**, Barplot showing the total number of cells that are immune versus non-immune (left) and cycling versus non-cycling (right) by MG status of tumor of origin. **f**, Boxplots comparing the cell type composition of bulk RNA seq samples after deconvolution using single cell RNA-seq signatures. Pairwise comparisons in each boxplot are statistically significant ($p < 0.05$), unless explicitly stated otherwise (ns, not significant). **g**, Heatmap showing the expression of marker genes for single cell clusters (determined by CIBERSORTx) in bulk RNA seq data. Each column represents one tumor. Rows are designated marker genes for each cluster. Tumors were partitioned into 4 partitions by consensus k-means clustering with samples and gene sets clustered by hierarchical clustering using Pearson distance metric.

Extended Data Fig. 11 Discrete and continuous patterns of variability can be identified in meningioma.

a, Hierarchical clustering of similarities between NMF programs. Top panel indicates Pearson correlations between number of mitochondrial and ribosomal genes detected with NMF scores for each program. A cluster of programs (dashed lines) showed positive correlation with the expression of

mitochondrial and ribosomal genes (confirmed by manual inspection). These programs were considered to be reflective of technical artifacts and not included in subsequent analyses. **b**, Violin plots showing the distribution of activation scores for NMF programs across MGs. **c**, Side-by-side tSNEs showing the relationship of discrete clustering results with activation scores of each NMF program. Shown are four representative samples. Activation scores of cell cycle program are closely associated with discrete clusters, whereas scores of metabolism, inflammatory, and mesenchymal program are not associated with discrete clusters. **d**, Heatmap showing the average expression of genes defining NMF programs (annotated to left) in representative sample CAM_0071. Cells are ranked and ordered according to the activation score of the metabolism program. There is a continuous pattern of gene expression variability in these programs.

Extended Data Fig. 12 Graphical summary of findings.

Shown is a schematic representation that summarizes the major molecular findings and conclusions of our study: unsupervised consensus clustering combining DNA copy number, DNA methylation, and mRNA sequencing data revealed four robust groups of tumors with prototypical biology and distinct clinical outcomes.

Supplementary information

Supplementary Information

This file contains the Supplementary Discussion, Supplementary References and Supplementary Figs. 1, 2.

Reporting Summary

Supplementary Tables

This file contains Supplementary Tables 1, 2, 3 and 6.

Supplementary Tables

This file contains Supplementary Tables 4, 5, 7 and 8.

Source data

Source Data Fig. 2

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Nassiri, F., Liu, J., Patil, V. *et al.* A clinically applicable integrative molecular classification of meningiomas. *Nature* **597**, 119–125 (2021). <https://doi.org/10.1038/s41586-021-03850-3>

- Received: 01 December 2020
- Accepted: 23 July 2021
- Published: 25 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03850-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03850-3>

- Article
- Open Access
- [Published: 04 August 2021](#)

The structural basis of odorant recognition in insect olfactory receptors

- [Josefina del Marmol¹](#),
- [Mackenzie A. Yedlin](#) ORCID: [orcid.org/0000-0003-4085-9509¹](https://orcid.org/0000-0003-4085-9509) &
- [Vanessa Ruta](#) ORCID: [orcid.org/0000-0002-4963-2768¹](https://orcid.org/0000-0002-4963-2768)

[Nature](#) volume **597**, pages 126–131 (2021)

- 11k Accesses
- 192 Altmetric
- [Metrics details](#)

Subjects

- [Cryoelectron microscopy](#)
- [Sensory processing](#)

Abstract

Olfactory systems must detect and discriminate amongst an enormous variety of odorants¹. To contend with this challenge, diverse species have converged on a common strategy in which odorant identity is encoded through the combinatorial activation of large families of olfactory receptors^{1,2,3}, thus allowing a finite number of receptors to detect a vast chemical world. Here we offer structural and mechanistic insight into how an individual olfactory receptor can flexibly recognize diverse odorants. We show that the olfactory receptor *MhOR5* from the jumping bristletail⁴ *Machilis hrabei* assembles as a homotetrameric odorant-gated ion channel with broad chemical tuning. Using cryo-electron microscopy, we elucidated the structure of *MhOR5* in multiple gating states, alone and in complex with two of its agonists—the odorant eugenol and the insect repellent DEET. Both ligands are recognized through

distributed hydrophobic interactions within the same geometrically simple binding pocket located in the transmembrane region of each subunit, suggesting a structural logic for the promiscuous chemical sensitivity of this receptor. Mutation of individual residues lining the binding pocket predictably altered the sensitivity of *MhOR5* to eugenol and DEET and broadly reconfigured the receptor's tuning. Together, our data support a model in which diverse odorants share the same structural determinants for binding, shedding light on the molecular recognition mechanisms that ultimately endow the olfactory system with its immense discriminatory capacity.

[Download PDF](#)

Main

The olfactory system faces a unique challenge amongst sensory modalities owing to the inordinate complexity of the chemical world. Whereas light waves vary continuously in amplitude and frequency, odorants differ discretely along an enormous number of dimensions in their molecular structure and physicochemical properties. Consequently, just three photoreceptors are sufficient to sense the entire spectrum of visible light, but large repertoires of olfactory receptors appear to be necessary to detect and discriminate amongst the diversity of chemicals in the environment^{1,2,3}. In mammals, odour detection is mediated by G-protein-coupled receptors that signal through canonical second-messenger cascades^{5,6}. By contrast, insects detect volatile chemicals using a unique class of odorant-gated ion channels^{7,8} consisting of two subunits: a conserved co-receptor (Orco) subunit^{9,10} and a highly divergent odorant receptor (OR) subunit that contains the odorant-binding site and confers chemical sensitivity to the heteromeric complex¹¹.

Although mammals and insects rely on distinct molecular mechanisms for odour detection, they share a common neural logic for olfactory perception based on the combinatorial activation of distinct ensembles of olfactory receptors and associated sensory neurons^{1,3,12}. Central to this sensory coding strategy is that most individual ORs detect a variety of structurally and chemically diverse odorants^{11,13,17,25}. However, in the absence of a structural model, how such flexible chemical recognition is achieved has remained unknown. Whether the broad chemical tuning of ORs reflects the presence of multiple odorant-binding sites that differ in their chemical specificity or a single promiscuous binding pocket is not known. Furthermore, which structural or chemical features of odorants are recognized by a receptor remains unclear. In this study, we leveraged the evolutionary diversification of insect ORs to elucidate the structures of a homomeric receptor from a basal insect species bound to different ligands. We reveal how a single receptor can detect a wide array of odorants through a single promiscuous binding site that recognizes the overall physicochemical

properties of each odorant rather than being tuned to any of their specific structural or molecular features, suggesting a structural basis for flexible chemical recognition.

MhOR5 is a broadly tuned receptor

Although neopteran insects each express a repertoire of highly divergent ORs along with a single, almost invariant Orco², recent genomic analyses have revealed that some basal insects, such as the jumping bristletail *M. hrabei*, possess only a small number of OR genes and lack an apparent Orco orthologue⁴ (Fig. 1a). *Mh*ORs have been proposed to represent the most ancestral members of the insect olfactory receptor family, arising before the emergence of Orco^{4,14}. Although little is known about chemosensory detection in the jumping bristletail, we reasoned that *Mh*ORs might function as homomeric olfactory receptors. We heterologously expressed each *Mh*OR in HEK293 cells and found that, indeed, *Mh*OR1 and *Mh*OR5 migrated as tetramers on non-denaturing native gels (Extended Data Fig. 1a,b). To assess whether these homomeric complexes function as chemoreceptors, we used a high-throughput fluorescence assay¹⁰ in which we co-expressed *Mh*OR1 or *Mh*OR5 with the indicator GCaMP6s and measured calcium influx in response to a panel of 54 small molecules over a range of concentrations. We found that *Mh*OR5 was activated by many volatile odorants but not tastants, consistent with a role for this receptor in olfactory detection (Fig. 1b, Extended Data Fig. 2a–d). *Mh*OR5 was also activated by the insect repellent DEET and inhibited by the synthetic Orco agonist VUAA1¹⁵. To quantitatively capture the complexity of odorant-evoked responses¹⁶ (Extended Data Fig. 2a–d), we defined an activity index for each odorant ($-\log(\text{EC}_{50}) \times \max \Delta F/F$; in which EC_{50} is the concentration of ligand at which the response reaches half its maximal value) that reflects both the apparent affinity and maximal efficacy of an odorant. *Mh*OR5 was activated by over 65% of odorants, resembling the broad molecular receptive fields of many insect and mammalian ORs^{11,13,17} (Extended Data Fig. 1d). By contrast, *Mh*OR1 exhibited far more selective tuning, responding to only eight odorants from the same chemical panel (Extended Data Fig. 1e). Both *Mh*OR1 and *Mh*OR5 were activated by ligands that spanned multiple chemical classes and a range of physicochemical properties (Extended Data Fig. 1e,f), exemplifying the complex chemical logic of odorant detection.

Fig. 1: The structure of *Mh*OR5, a broadly tuned odorant-gated ion channel.

 **figure1**

a, Phylogenetic tree of selected insect clades and the representative numbers of OR and Orco genes in their genomes. Myr, million years. **b**, Left, activity of *MhOR5* evoked by a panel of 54 small molecule ligands. Right, dose–response curves of *MhOR5* to selected ligands. For dose–response curves for all ligands, see Extended Data Fig. 2; *n* values for biological replicates are in Supplementary Table 1. See [Methods](#) for a detailed description of the activity index. **c**, Representative traces of eugenol-evoked currents in HEK cells expressing *MhOR5*. Top, whole-cell currents at voltages from −80 mV to 40 mV. Bottom, single-channel recordings in outside-out patches at −80 mV. Red dashed lines indicate current levels when different numbers of channels open. **d**, Cryo-EM structure of *MhOR5* shown from the side (left) and top (right). Subunits are coloured in rainbow palette from the N terminus (purple) to the C terminus (red). In the side view, front and back subunits were removed for clarity of visualization. Black lines, membrane boundaries.

Whole-cell recordings of HEK cells expressing *MhOR5* showed that the odorant eugenol elicited slowly activating inward currents that reversed at 0 mV, consistent with its function as a cation-selective ion channel (Fig. 1c). In outside-out patches, eugenol evoked small-conductance single-channel activity that rapidly flickered between the closed and open states, resembling canonical heteromeric insect olfactory receptors^{7,10} (Fig. 1c, Extended Data Fig. 2e). *MhORs* thus autonomously assemble as homotetrameric odorant-gated ion channels and display the divergent chemical tuning profiles typical of this receptor family. Given its sensitivity to a broad array of structurally diverse odorants, we focused on *MhOR5* to investigate the molecular basis of promiscuous chemical recognition.

Structure of the *MhOR5* homotetramer

We used single-particle cryo-electron microscopy (cryo-EM) to elucidate the structure of the *MhOR5* tetramer. We obtained a density map at 3.3 Å resolution (Fig. 1d, Extended Data Figs. 3, 4, Extended Data Table 1), which allowed us to unambiguously build a model for the majority of the protein, with the exception of several extra-membranous loops and the short intracellular N terminus and extracellular C terminus (Extended Data Figs. 4c, 5c). A comparison of the structure of *MhOR5* with the previously elucidated structure of Orco from the wasp *Apocrypta bakeri*¹⁰ showed that these two receptors, despite sharing only about 18% amino acid conservation, display notable similarity, both in the fold of each heptahelical subunit and in the tetrameric organization of the subunits within the membrane plane (Extended Data Fig. 5a, b). As in Orco, each *MhOR5* subunit contributes a single helix (S7b) to the central ion conduction pathway, and their S0–S6 helices form a loosely packed domain that projects radially away from the pore axis (Fig. 1d). Within the membrane, the contacts between *MhOR5* subunits are minimal and confined to the pore, whereas about 75% of the residues that form inter-subunit interactions reside within the intracellular ‘anchor’ domain, formed from the intertwined S4–S7 helices of all four subunits (Extended Data Fig. 5d). Analogous to the Orco structure, the tightly packed anchor domain of *MhOR5* exhibited the highest local resolution (Extended Data Fig. 4c), consistent with a structural role in stabilizing the loosely assembled S0–S6 transmembrane domains within the lipid bilayer. The limited sequence conservation across neopteran ORs and Orcos maps to residues predominantly within the pore and anchor domain¹⁰, further underscoring how the architecture of this receptor family can accommodate a high degree of sequence diversification while maintaining the same overall fold, a feature that is likely to have facilitated the rapid evolution of ORs.

Odorant binding leads to pore opening

To explore the structural determinants of odorant gating, we determined a 2.9 Å-resolution structure of *MhOR5* in complex with its highest activity ligand, eugenol. Three-dimensional reconstruction of the bound structure immediately yielded higher resolution, as was apparent from early stages of data processing (Extended Data Figs. 3, 4, Extended Data Table 1). The *MhOR5* pore displays the same distinct quadrivial architecture as the Orco homotetramer¹⁰, in which a single extracellular pathway opens into a large aqueous vestibule near the intracellular surface of the membrane and then diverges into four lateral conduits formed at the interfaces between subunits (Fig. 2a, b). In the apo structure, the S7b helices coalesce to form the narrowest portion of the ion conduction pathway. In particular, Val468 protrudes into the channel lumen, generating a hydrophobic constriction measuring about 5.3 Å in diameter, and thus serves as a gate to impede the flow of hydrated ions through the quadrivial pore (Fig.

[2a, b, d](#)). In the presence of eugenol, the extracellular aperture of the pore is dilated as a result of movement of the S7b helices away from the central pore axis (Fig. [2b-d](#)), which rotates Val468 out of the pore lumen to face the lipid bilayer, while Gln467 rotates in to face the ion pathway. As a consequence of this rearrangement, the chemical environment of the pore is transformed from a narrow hydrophobic constriction to a wide hydrophilic ring, 9.2 Å in diameter, that can readily accommodate the passage of hydrated cations. Notably, the remainder of the quadrivial pore remains essentially unaltered with the addition of eugenol (Fig. [2a-c](#)), confirming that the tightly packed anchor domain forms a relatively stationary structural element¹⁰. The dilation of the S7b helices thus appears to be sufficient to gate the ion conduction pathway—this small conformational change would present a low energetic barrier to gating, consistent with the low affinity of most odorants^{11,17} and with functional evidence that *MhOR5* channels, as with many insect olfactory receptors, open spontaneously even in the absence of ligand^{7,11} (Extended Data Fig. [1c](#)).

Fig. 2: Odorant-evoked opening of *MhOR5*.



a, c, The channel pores of the apo (**a**, blue) and eugenol-bound *MhOR5* (**c**, pink). Black dashed lines, membrane boundaries. **b**, The diameter of the ion conduction pathway (solid lines) and along the anchor domain (dashed lines). *y*-axis shows distance from the outer membrane boundary towards the intracellular space in Å. **d**, Close-up view of the pore from the extracellular side in the apo (left, blue) and eugenol-bound (right, pink) structures. Distances between S7b residues measured from atom centres. **e–g**, Dose–response curves (left) and mean ± s.e.m. log(EC₅₀) (right) for wild-type (WT), Val468 mutants and Gln467 mutants of *MhOR5* (**e**), wild-type and Q472A mutant (homologous to Gln467 in *MhOR5*) of Orco as a homotetramer (**f**) and wild-type and Q472A mutant of Orco in heteromeric Orco–AgOR28 complex (**g**). Statistical significance determined using one-way ANOVA followed by Dunnett’s multiple comparison tests. For mutants for which the EC₅₀ was incalculably high and Bartlett’s test showed non-homogenous variance, statistical significance determined with a Brown–Forsythe test. ****P < 0.0001; ***P < 0.001; *P < 0.01; NS, not significant. Supplementary Tables 2, 3 contain details including *n* values for all biological replicates.

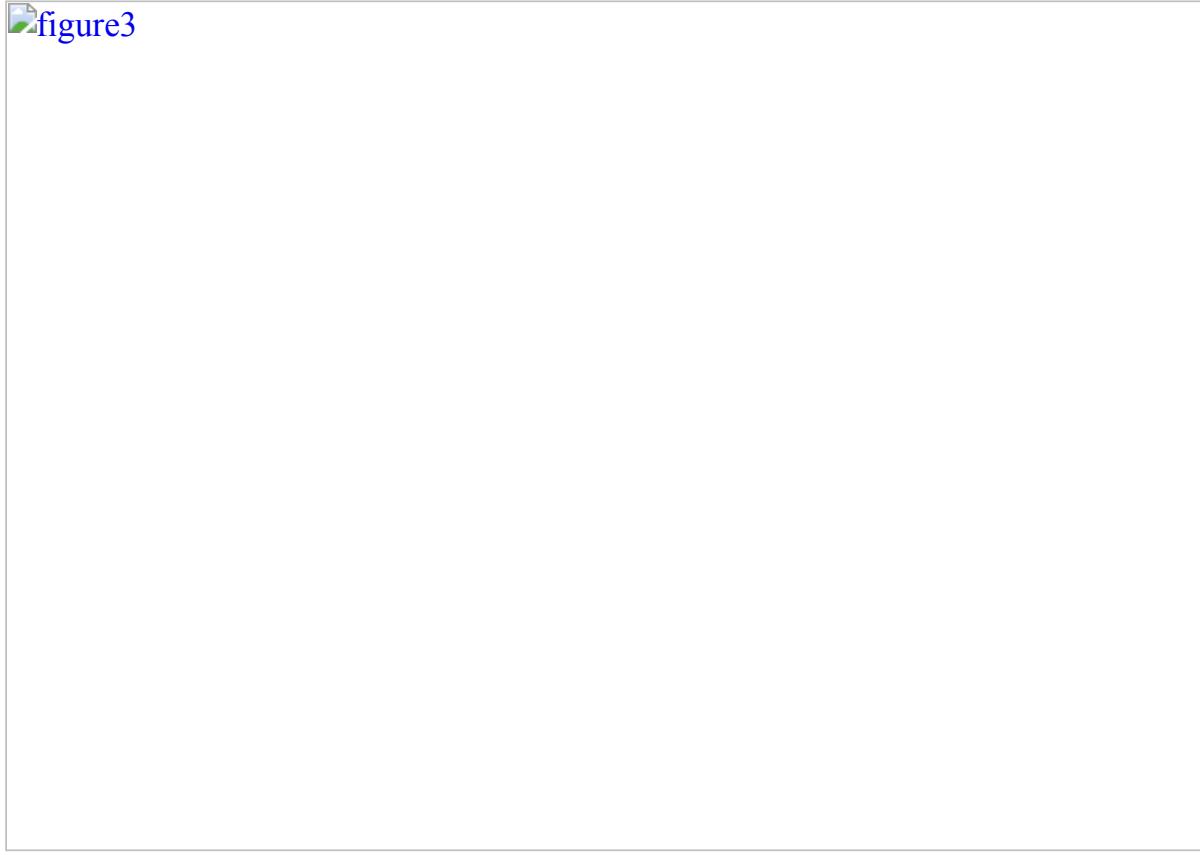
Gln467 is highly conserved across Orcos and ORs from *M. hrabei* and other basal insect species¹⁴ and was previously identified as a component of the only signature sequence motif (TYhhhhhQF, in which h is any hydrophobic amino acid) that is diagnostic of the highly divergent insect chemosensory receptor superfamily¹⁸. Mutation of Gln467 in *MhOR5* to either the smaller residue alanine or the positively charged residue arginine strongly impaired receptor function, whereas a more conservative mutation to asparagine had little effect on activity (Fig. 2e). Replacement of the neighbouring residue Val468 with either alanine or glutamine resulted in minimal changes to odorant activation (Fig. 2e), highlighting that movement of the pore helices, rather than simply the presence of a pore-lining glutamine, is necessary to gate the channel. In the closed structure of the Orco homotetramer¹⁰, the homologous residue, Gln472, points into the lipid membrane, similar to its position in the closed conformation of *MhOR5*. Mutation of Gln472 to alanine in Orco yielded non-functional homomeric channels (Fig. 2f). Gln472 is thus one of the few S7b residues in Orco that is intolerant to such a perturbation¹⁰, consistent with a conserved and critical role for this residue in gating and/or ion permeation across this receptor family. Notably, the Q472A Orco mutant could be partially rescued by co-expression with an OR from *Anopheles gambiae* (Fig. 2g), indicating that this mutant can fold and function in the context of the heteromeric assembly and underscoring the intrinsic robustness of the Orco–OR complex, where both subunits contribute to a shared ion conduction pathway^{10,19}.

Architecture of the odorant-binding site

In the transmembrane domain of each *MhOR5* subunit, the S2, S3, S4 and S6 helices splay apart to form a 15 Å-deep pocket within the extracellular leaflet of the bilayer (Fig. 3a, Extended Data Figs. 6, 7). Clearly defined density consistent with the size and shape of eugenol lies at the base of this pocket, enclosed within a hydrophobic box constructed from several large aromatic and hydrophobic residues, with Trp158 forming the lid, Tyr91 and Tyr383 forming its base, and flanked by Tyr380 on one side and by Met209, Ile213 and Phe92 on the other (Fig. 3b, Extended Data Figs. 6, 9d). In the apo structure, the density for some of these amino acids was diffuse (Extended Data Fig. 6b), which could be attributed to the overall lower resolution of this structure or to conformational flexibility when no odorant is bound. The lower resolution of the apo pocket precluded us from defining the path that eugenol takes to enter the pocket, as in the bound structure the pocket is not obviously accessible to solvent (Extended Data Figs. 6b, 7a), or from determining whether the cavity is partially occupied in the absence of an added ligand. Binding of eugenol, however, stabilized the residues that line the pocket, allowing unambiguous mapping of the side chains that form the binding site.

Fig. 3: Architecture of the odorant-binding site in *MhOR5*.

figure3



a, Left, side view of two subunits of *MhOR5* shown as a surface representation with the cross-section through the binding pocket indicated. Right, top view of a cross-

section of the *MhOR5* tetramer through the binding pocket. Bottom, expanded view of a single subunit shown as a surface representation with helices coloured and labelled. Eugenol shown in stick representation. **b**, View of the binding pocket. Pink, residues in contact with eugenol; yellow, eugenol; black mesh, cryo-EM density for eugenol. **c**, Mutagenesis of residues in contact with eugenol (gold) and two neighbouring residues (purple) that project away from the pocket. Mean \pm s.e.m. $\log(\text{EC}_{50})$ shown. Statistical significance determined using one-way ANOVA followed by Dunnett's multiple comparison tests. For mutants for which the EC_{50} was incalculably high and Bartlett's test showed non-homogenous variance, statistical significance determined with a Brown–Forsythe test. *** $P < 0.0001$; * $P < 0.05$; NS, not significant. Dose–response curves shown in Extended Data Fig. 9, additional views of the pocket in Extended Data Figs. 6, 8, 9d, and n values for all biological replicates in Supplementary Table 2.

To explore the potential binding modes of eugenol, we used computational docking methods²⁰ and performed a broad grid search spanning the majority of the transmembrane domain. This analysis identified a series of closely related eugenol poses with uniformly favourable docking scores that fit into the experimental density well (Extended Data Fig. 8a). At this resolution, differentiating between these poses is challenging given that eugenol, as with most odorants, is a small molecule with few distinguishing structural features to orient it within the density. Nevertheless, eugenol was predicted to bind through comparable interactions across all the top poses, but these interactions could be mediated by different hydrophobic or aromatic residues within the pocket. For example, the benzene ring of eugenol was stabilized through π -stacking interactions, but these could be mediated by Trp158, Tyr91, or Tyr380, which lie on opposing faces of the binding pocket. In every pose, eugenol also formed extensive hydrophobic interactions with an overlapping complement of aliphatic and aromatic side chains. Moreover, although eugenol's hydroxyl was consistently oriented towards the only polar amino acid lining the pocket (Ser151), none of the predicted poses adopted a geometry that allowed them to form hydrogen bonds with the surrounding residues. Therefore, recognition of eugenol appears to rely on non-directional hydrophobic interactions formed with a distributed array of binding pocket residues. Although only one of these poses might be energetically favoured, structural studies of odorant binding proteins^{21,22} that ferry hydrophobic ligands through the sensory neuron lymph have revealed that an individual odorant can bind in different poses within the same hydrophobic binding cavity; thus, it is possible that eugenol might likewise sample from multiple energetically degenerate binding modes in *MhOR5*.

To functionally corroborate the eugenol binding site, we identified ten amino acids with side chains that were in close proximity to the ligand density—Val88, Tyr91, Phe92, Ser151, Gly154, Trp158, Met209, Ile213, Tyr380 and Tyr383—and found that mutation of any of these residues to alanine strongly altered eugenol signalling (Fig.

[3c](#), Extended Data Fig. [9a–c](#)). Several of these mutants also displayed increased baseline activity (Extended Data Fig. [9a, e](#)), suggesting that these residues stabilize the closed conformation. Mutation of adjacent residues that project away from the binding site—Thr87 and Leu379—had minimal effect on activation by eugenol (Fig. [3c](#), Extended Data Figs. [6a, 9a](#)), underscoring the specificity of these perturbations to odorant-dependent gating.

A comparison of the apo and eugenol-bound structures indicates that, in addition to the dilation of the pore, smaller conformational changes appear to be distributed throughout the transmembrane portion of the S0–S6 helices (Extended Data Fig. [10a, b](#) and Supplementary Videos [1, 2](#)). Although the delocalized nature of these small rearrangements makes it challenging to delineate how odorant binding is transduced to pore opening, one potential route is through the S5 helix, which runs parallel to the S7b helix that lines the pore and anti-parallel to the S6 helix that contributes key residues to the odorant-binding pocket. Upon eugenol binding, these three helices move together away from the central axis of the channel, displacing the S7b helices outwards to gate the ion conduction pathway (Extended Data Fig. [10a, b](#)). Close to the extracellular surface of the membrane, the S5 and S7 helices interact through Tyr362 and Leu465, which are highly conserved as hydrophobic amino acids and evolutionarily coupled²³, pointing to a coordinated role in receptor function. These residues remain tightly packed as the S7b helix moves into an open configuration (Extended Data Fig. [10b](#)), suggesting that they might couple conformational rearrangements within the odorant-binding pocket to the dilation of the pore. Mutation of either Tyr362 or Leu465 to alanine impaired eugenol activation, whereas mutation of Tyr362 to phenylalanine had no effect (Extended Data Fig. [10c](#)), supporting a model in which hydrophobic interactions at this position contribute to gating.

Structural basis of odorant specificity

To investigate the diversity of binding modes used by different ligands, we determined the 2.9 Å structure of *MhOR5* in complex with the insect repellent DEET (Extended Data Table [1](#)). The S7b helices in the DEET-bound structure were dilated to a diameter of 8.7 Å (Extended Data Fig. [10d–f](#)), indicating that different ligands elicit a common conformational change to gate the pore. Density corresponding to DEET localized to the same binding pocket as eugenol, encased within the same box-like configuration of aromatic and aliphatic side chains (Fig. [4a, b](#), Extended Data Figs. [6b, 9d](#)). As with eugenol, computational docking of DEET yielded multiple poses with comparable docking scores that fit the experimental density well (Extended Data Fig. [8a](#)). Whereas each of the top poses was predicted to adopt a distinct orientation, all were stabilized through a similar complement of hydrophobic and/or π-stacking interactions. Although we cannot determine whether DEET adopts only one or multiple conformations within the binding pocket, these observations reinforce how non-directional hydrophobic

interactions may contribute to flexible chemical recognition, allowing different ligands to bind to the same structural locus or potentially enabling a single odorant to sample from multiple poses within the binding cavity.

Fig. 4: Structure-based mutagenesis retunes *MhOR5*.

 figure4



a, Cross-section of the binding pocket of an *MhOR5* subunit shown as a surface representation in complex with DEET (shown in stick representation within the pocket). Helices are shown coloured and labelled. **b**, View of the binding pocket (orientation as in Fig. 3b). Black mesh, cryo-EM density for DEET. **c**, Top, overlay of the *MhOR5* binding pocket in complex with DEET (teal) and in complex with eugenol (pink). Bottom, effect of mutating Met209 and Ile213 into residues with different length side-chains on eugenol and DEET signalling. Mean \pm s.e.m. $\log(\text{EC}_{50})$ shown.

Statistical significance determined using one-way ANOVAs followed by Dunnett's multiple comparison tests comparing mutants to their respective WT controls for each ligand. For mutants for which the EC₅₀ was incalculably high and Bartlett's test showed non-homogenous variance, statistical significance determined with a Brown–Forsythe test. ***P < 0.0001; NS, not significant. **d**, Tuning curves of *MhOR5* M209V (top) and I213M (bottom) mutants in response to a panel of 40 odorants, overlaid and ordered by high-to-low activity index in wild-type *MhOR5*. Additional dose–response curves shown in Extended Data Fig. 9, additional views of the pocket in Extended Data Figs. 6, 9d, and n values for all biological replicates in Supplementary Tables 2, 4, 5.

To investigate whether the broader panel of *MhOR5* ligands is recognized through a similar structural logic, we examined how their physicochemical descriptors correlated with receptor activity. Multiple regression analysis revealed that although no single metric was strongly predictive of agonism, the descriptors that best accounted for *MhOR5* activity were low polar surface area, low water solubility, and low potential for forming hydrogen bonds (Extended Data Table 2), consistent with our structural observations of a geometrically simple binding site in which diffuse hydrophobic interactions dominate. *MhOR1* agonism was less correlated with these descriptors, suggesting that they have a heterogeneous role in shaping the tuning of different receptors (Extended Data Table 2). Furthermore, the top 31 *MhOR5* agonists identified in our panel could be docked within this same binding site with favourable scores, stabilized predominantly through hydrophobic interactions (Extended Data Fig. 8), suggesting that diverse odorants can be recognized through distributed and non-directional interactions with an overlapping subset of residues in the *MhOR5* binding pocket.

A comparison of the eugenol and DEET-bound structures reveals how the *MhOR5* binding pocket might accommodate such diverse ligands. The constellation of amino acids lining the binding pocket retains the same overall geometry in both structures, leaving the architecture of the hydrophobic box largely unchanged. However, a small displacement of the S4 helix results in expansion of the pocket, probably to accommodate the longer aliphatic moiety of DEET and avoid a steric clash with the side chain of Met209 (Fig. 4c, Extended Data Fig. 6b). Functional data support these structural observations. Mutation of Met209 to smaller hydrophobic amino acids (valine or alanine) enhanced the affinity of DEET (Fig. 4c, Extended Data Fig. 9b). The same mutations attenuated eugenol sensitivity, suggesting that this smaller odorant occupies the binding pocket less optimally in the absence of the bulky methionine side chain. Conversely, mutation of Ile213, another aliphatic S4 residue that lies in close proximity to DEET, to the larger residue methionine abolished DEET sensitivity but marginally altered eugenol signalling (Fig. 4c, Extended Data Fig. 9c). Structure-guided mutagenesis therefore differentially altered the sensitivity of *MhOR5*.

to these two ligands. Furthermore, the I213M and M209V mutations broadly reconfigured the tuning of *MhOR5* to a larger panel of 40 odorants (Fig. 4d), supporting a model in which diverse chemicals are recognized by shared structural elements within a common binding pocket. Changes in odorant tuning, however, did not adhere to a simple logic, consistent with the complexity of physicochemical properties that define *MhOR5* agonism (Extended Data Table 2) and with the proposal that both the global geometry and local chemical environment of the binding pocket contribute to its chemical sensitivity.

To assess whether *MhOR5* can serve as a structural model for chemical recognition in other ORs, we used sequence homology to identify ten residues predicted to line the binding pocket in the more narrowly tuned *MhOR1* and examined their contribution to odorant tuning (Extended Data Fig. 11a–c). For all but one of these residues, mutation to alanine impaired *MhOR1* activation by its ligands, 1-octanol and eugenol, indicating that the odorant binding pocket is a conserved structural feature of this family, even between divergent receptors that display distinct chemical tuning. Furthermore, mutation of Met231 in *MhOR1* to the corresponding residue in *MhOR5*, isoleucine, enhanced the sensitivity of *MhOR1* to a panel of odorants (Extended Data Fig. 11d). Thus, whereas the I213M mutation narrows the chemical tuning of *MhOR5*, the reciprocal M231I mutation broadens the molecular receptive range of *MhOR1*, shifts in sensitivity that could be attributed to alterations in the size of the binding pocket. Odorant recognition in different insect olfactory receptors appears therefore to rely on a conserved binding site that can be readily retuned to detect different regions of chemical space.

Discussion

The broad tuning of olfactory receptors is central to the detection and discrimination of the vast chemical world. Here we show that *MhOR5* detects a wide array of odorants through a single promiscuous binding site, offering structural insight into how such flexible chemical recognition is achieved. Notably, odorant binding relies predominantly on hydrophobic interactions, which lack the strict geometric constraints inherent to other intermolecular associations (such as hydrogen bonds) that frequently mediate ligand recognition. The distributed arrangement of hydrophobic and aromatic residues across multiple surfaces of the binding pocket further relaxes orientational constraints by allowing odorants to form comparable interactions with many of its faces. Moreover, the simple geometry of the binding site imposes minimal restriction on the shape of odorants that can bind, accommodating both eugenol and DEET with little structural rearrangement. Computational docking analyses support these structural observations and suggest that the same logic underlies the sensitivity of *MhOR5* to structurally and chemically diverse ligands. The prevalence of comparatively weak intermolecular interactions is compatible with the low affinity of

most odorants^{11,13,17,24} and the small conformational change required to gate the channel. Olfactory receptor tuning thus depends on the stereochemistry of its ligands^{25,26}, but does not adhere to the classic lock-and-key mechanism that governs many receptor–ligand interactions.

Residues that have been implicated in odorant specificity in different neopteran receptors map to the binding pocket of *MhOR5*^{10,27,28,29,30}, indicating that it represents a conserved and canonical locus for odorant detection across this highly divergent family. Binding of DEET to the same site offers structural corroboration that this insect repellent might exploit the promiscuity of diverse ORs and serve as a molecular ‘confusant’ by scrambling the olfactory code³¹. Other modulators of olfactory receptors, such as VUAA1 (which inhibits *MhOR5*), cannot favourably dock within this binding pocket owing to their much larger size, suggesting that insect olfactory receptors might possess additional points of allosteric modulation that expand their signalling mechanisms.

Several important implications arise from our observation that diverse odorants share the same structural determinants for binding. Notably, even single conservative mutations within the binding pocket can broadly reconfigure the chemical tuning of the receptor, a feature that is likely to have facilitated the rapid evolution of receptors with distinct ligand specificity^{2,27,28,29}. However, such extensive retuning also poses a substantial evolutionary constraint, as individual binding-site mutations are likely to have a pleiotropic effect on the representation of multiple odorants, potentially serving to broadly reconfigure the odour code. The promiscuous and arbitrary nature of odorant recognition is likely to impose substantial selective pressures on the structure and function of olfactory circuits, driving the evolution of synaptic and circuit mechanisms that can decorrelate, decode, and impose meaning onto combinatorial patterns of receptor activity³². Odour discrimination is thus transformed from a biochemical problem at the receptor level to a neural coding problem within the brain.

Although the structure of a mammalian olfactory receptor has yet to be elucidated, odorant detection in mammals has been proposed to also rely on distributed hydrophobic and non-directional interactions within a deep transmembrane pocket^{33,34,35}. Structurally and mechanistically distinct receptor families appear to therefore rely on similar principles for their broad chemical tuning, pointing to common constraints in how diverse hydrophobic molecules can be recognized. Additional mechanisms for odorant recognition certainly exist, in particular for receptors that are selectively tuned to ethologically relevant chemical classes, such as pheromones³⁶, the perceptual meaning of which is singular and invariant. Whether stricter odorant specificity relies on distinct intermolecular binding modes, variations in the geometry of the binding pocket, or both, remains to be determined.

Finally, our work sheds light on the evolution of the insect olfactory system. We demonstrate that *Mh*ORs can function as homomeric odorant-gated channels, supporting the proposal that they lie at the ancestral origin of the insect olfactory receptor family^{4,14}, which expanded massively across insect lineages to emerge as possibly the largest and most divergent class of ion channels in nature². Why neopteran ORs became obligate heteromers with Orco remains unclear, but presumably reflects the fact that Orco confers structural stability on the complex, thereby relaxing evolutionary constraints on the ORs and allowing them to further diversify, to ultimately support the flexible detection and discrimination of an enormous and ever-changing chemical world.

Methods

Expression and purification of *Mh*OR5

The coding sequence of *M. hrabei* OR5 (*Mh*OR5) was synthesized as a gene fragment (Twist). Residues Lys2 to Pro474 were cloned into a pEG BacMam vector³⁷ containing N-terminal tags of Strep II, superfolder GFP³⁸, and an HRV 3C protease site for cleavage (N-CACCatg-ST2-SGR-sfGFP-PPX-Ascl-*Mh*OR5-taa-NotI-C). The Ascl/NotI restriction enzyme sites enable efficient cloning of different OR sequences. SF9 cells (ATCC CRL-1711) were used to produce baculovirus containing the *Mh*OR5 construct, and the virus, after three rounds of amplification, was used to infect HEK293S GnTI⁻ cells (ATCC CRL-3022)³⁷. Cell lines were not authenticated except as performed by the vendor. HEK293S GnTI⁻ cells were grown at 37 °C with 8% carbon dioxide in Freestyle 293 medium (Gibco) supplemented with 2% (v/v) fetal bovine serum (Gibco). Cells were grown to 3×10^6 cells/ml and infected at a multiplicity of infection of about 1. After 8–12 h, 10 mM sodium butyrate (Sigma-Aldrich) was added to the cells and the temperature was dropped from 37 °C to 30 °C for the remainder of the incubation. Seventy-two hours after initial infection, cells were collected by centrifugation, washed with phosphate-buffered saline (pH 7.5; Gibco), weighed and flash frozen in liquid nitrogen. Pellets were stored at –80 °C until they were thawed for purification.

For purification, cell pellets were thawed on ice and resuspended in 20 ml lysis buffer per gram of cells. Lysis buffer was composed of 50 mM HEPES/NaOH (pH 7.5), 375 mM NaCl, 1 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µg/ml pepstatin A, 1 mM phenylmethylsulfonyl fluoride (PMSF; all from Sigma-Aldrich) and about 3 mg DNase I (Roche). *Mh*OR5 was extracted using 0.5% (w/v) *n*-dodecyl β-d-maltoside (DDM; Anatrace) with 0.1% (w/v) cholesterol hemisuccinate (CHS; Sigma-Aldrich) for 2 h at 4 °C. The mixture was clarified by centrifugation at 90,000g and the supernatant was added to 0.1 ml StrepTactin Sepharose resin (GE Healthcare) per

gram of cells and rotated at 4 °C for 2 h. The resin was collected and washed with 10 column volumes (CV) of 20 mM HEPES/NaOH, 150 mM NaCl with 0.025% (w/v) DDM and 0.005% (w/v) CHS (together, SEC buffer). *MhOR5* was eluted by adding 2.5 mM desthiobiotin (DTB) and cleaved overnight at 4 °C with HRV 3C Protease (EMD Millipore). Sample was concentrated to about 5 mg/ml and injected onto a Superose 6 Increase column (GE Healthcare) equilibrated in SEC buffer. Peak fractions containing *MhOR5* were concentrated until the absorbance at 280 nm reached 5–6 (approximately 5 mg/ml) and immediately used for grid preparation and data acquisition. For the eugenol-bound structure, peak fractions were pooled, and eugenol (Sigma Aldrich, CAS#97-53-0) dissolved in dimethylsulfoxide (DMSO; both Sigma-Aldrich) was added for a final odour concentration of 0.5 mM, and the complex was incubated at 4 °C for 1 h. The maximum DMSO concentration was kept below 0.07%. The complex was then concentrated to approximately 5 mg/ml and used for grid preparation. For the DEET-bound structure, sample from the overnight cleavage step was concentrated to about 5 mg/ml and injected into the Superose 6 Increase column equilibrated in SEC buffer with 1 mM DEET (Sigma Aldrich, CAS#134-62-3). Peak fractions were concentrated to about 5 mg/ml and used immediately for grid preparation.

Cryo-EM sample preparation and data acquisition

Cryo-EM grids were frozen using a Vitrobot Mark IV (FEI) as follows: 3 µl of the concentrated sample was applied to a glow-discharged Quantifoil R1.2/1.3 holey carbon 400 mesh gold grid, blotted for 3–4 s in >90% humidity at room temperature, and plunge frozen in liquid ethane cooled by liquid nitrogen.

Cryo-EM data were recorded on a Titan Krios (FEI) operated at 300 kV, equipped with a Gatan K2 Summit camera. SerialEM³⁹ was used for automated data collection. Movies were collected at a nominal magnification of 29,000× in super-resolution mode resulting in a calibrated pixel size of 0.51 Å/pixel, with a defocus range of approximately −1.0 to −3.0 µm. Fifty frames were recorded over 10 s of exposure at a dose rate of 1.22 electrons per Å² per frame.

Movie frames were aligned and binned over 2 × 2 pixels using MotionCor2⁴⁰ implemented in Relion 3.0⁴¹, and the contrast transfer function parameters for each motion-corrected image were estimated using CTFFIND4⁴².

Apo structure

Two datasets were collected with 4,050 micrographs in dataset A and 3,748 micrographs in dataset B. Processing was done independently for each dataset in the following way: particles were picked using a 3D template generated in an initial model

from a dataset of 5,000 particles picked in manual mode. A total of 562,794 (dataset A) and 536,145 (dataset B) particles were subjected to 2D classification using RELION-3.0⁴¹. Particles from the best 2D classes (210,833 for dataset A, 183,061 for dataset B) were selected and subjected to 3D classification imposing C4 symmetry and adding a soft mask to exclude the detergent micelle after 25 iterations. One class from each dataset containing 44,884 (dataset A) and 43,788 (dataset B) particles was clearly superior in completeness and definition of the transmembrane domains. These particles were subjected to 3D refinement with C4 symmetry, followed by Bayesian polishing and CTF refinement. The polished particles from both datasets were exported to cryoSPARC v2⁴³ and processing continued with the joined dataset of 88,672 particles. In cryoSPARC, further heterogeneous refinement resulted in a single class with 49,832 particles that were subjected to particle subtraction with a micelle mask. Non-uniformed refinement of subtracted particles imposing C4 symmetry yielded the final map with an overall resolution of 3.3 Å as estimated by cryoSPARC with a cutoff for the Fourier shell correlation (FSC) of 0.143⁴⁴.

Ligand-bound structures

Processing for the eugenol-bound and DEET-bound structures occurred through the following pipeline: 4,410 (eugenol) and 4,365 (DEET) micrographs were collected and used to pick 461,254 (eugenol) and 787,448 (DEET) particles that were extracted, unbinned and exported into cryoSPARC v2. In cryoSPARC, several rounds of 2D classification resulted in 221,339 (eugenol) and 180,874 (DEET) particles that were used to generate an initial model with four classes with no imposed symmetry. These models were inputted as templates of a heterogeneous refinement with no imposed symmetry, from which one (eugenol) and two (DEET) final classes were selected containing 129,031 (eugenol) and 121,441 (DEET) particles. These particles were refined and exported to RELION 3.0 where they were subjected to a round of 3D classification with no imposed symmetry. The best class from this 3D classification contained 54,900 (eugenol) and 56,191 (DEET) particles that were subjected to Bayesian polishing and CTF refinement. Polished particles were then imported into cryoSPARC v2 and subjected to particle subtraction. Final non-uniform refinement with C4 symmetry imposed resulted in the final maps with overall resolution of 2.9 Å in both cases, estimated with a cutoff for the FSC of 0.143. In all cases, the four-fold symmetry of the channel was evident from the initial 2D classes without having imposed symmetry and refinements without imposed symmetry produced four-fold symmetric maps.

Model building

The Cryo-EM structure of Orco (Protein Data Bank (PDB) accession 6C70) was used as a template for homology modelling of *MhOR5* using Modeller⁴⁵, followed by manual building in Coot⁴⁶. The 3.3 Å density map of the apo was of sufficient quality to build the majority of the protein, with the exception of the S3–S4 and S4–S5 loops, the 13 N-terminal residues and the 5 C-terminal residues. The models were refined using real-space refinement implemented in PHENIX⁴⁷ for five macro-cycles with four-fold non-crystallographic symmetry applied and secondary structure restraints applied. The eugenol- and DEET-bound models were refined including the ligands, which were placed as a starting point within the corresponding density in a pose that was obtained through docking methods (described below) and with restraints obtained with the electronic Ligand Builder and Optimization Workbench⁵⁸ (eLBOW) implemented in PHENIX. Model statistics were obtained using MolProbity. Models were validated by randomly displacing the atoms in the original model by 0.5 Å, and refining the resulting model against half maps and full map⁴⁸. Model–map correlations were determined using phenix.mtriage. Images of the maps were created using UCSF ChimeraX⁴⁹. Images of the model were created using PyMOL⁵⁰ and UCSF ChimeraX⁴⁹.

Docking analysis

All compounds were docked using Glide^{20,51} implemented in Maestro (Schrödinger, suite 2020). In brief, the model was imported into Maestro and prepared for docking. A 20 Å³ cubical grid search was built centred in the region of observed ligand density. Ligand structures were imported into Maestro by their SMILES unique identifiers and prepared using Epik⁵² to generate their possible tautomeric and ionization states, all optimized at pH 7.0 ± 2. All ligands were docked within the built grid, and the top poses that best fit the density are presented in Extended Data Fig. 8. The top activators scored with values between −7.4 and −4. While all activators docked with negative scores, some non-activators also docked with favourable scores. For example, caffeine docked favourably despite the molecule not activating the channel in our functional experiments. As docking does not incorporate dynamics of the receptor, it is not expected that docking will correlate homogeneously or monotonically with experimentally determined activity of ligands. At most a qualitative agreement can be expected.

Structure analysis

Residues at subunit interfaces were identified using PyMOL as any residue within 5 Å of a neighbouring subunit (Extended Data Fig. 5d). The pore diameters along the central axis and lateral conduits were calculated using the program HOLE⁵³, which models atoms as solid spheres of Van der Waals radius (Fig. 2a–c, Extended Data Fig.

[10d, e](#)). Two calculations were performed for each structure: one along the central four-fold axis (central pore) and another between subunits near the cytosolic membrane interface (lateral conduits). The plots in Fig. [2b](#) and Extended Data Fig. [10e](#) show the diameter along the central axis of the main conduit and the lateral conduit. The measurements in Fig. [2d](#) and Extended Data Fig. [10f](#) between residues lining the pore are taken from atom centres using PyMol. Electrostatic surface representations were performed using ChimeraX v1.1, coulombic estimation with default parameters (Extended Data Fig. [7](#)). Morph videos were created in ChimeraX v1.1 with direct interpolation between states.

Electrophysiology

HEK293 cells were maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) GlutaMAX (all Gibco) at 37 °C with 5% (v/v) carbon dioxide. Cells were plated on 35-mm tissue-culture-treated Petri dishes 72–48 h before recording, and infected with the same pEG BacMam GFP-tagged *MhOR5* construct used for expression 24–48 h before recording. Electrodes were drawn from borosilicate patch glass (Sutter Instruments) and polished (MF-83, Narishige Co.) to a resistance of 3–6 MΩ when filled with pipette solution. Analogue signals were digitized at 20 kHz (Digidata 1440A, Molecular Devices) and filtered at 1 kHz (whole-cell) or 2 kHz (patch recordings) using the built-in four-pole Bessel filter of a Multiclamp 700B patch-clamp amplifier (Molecular Devices) in whole-cell or patch mode. Whole-cell recordings were baseline-subtracted offline. Patch signals were further resampled offline for representations.

Whole-cell and single-channel recordings in Fig. [1c](#) and Extended Data Fig. [2e](#) were performed using an extracellular (bath) solution composed of 135 mM NaCl, 5 mM KCl, 2 mM MgCl₂, 2 mM CaCl₂, 10 mM glucose, 10 mM HEPES-Na/HCl (pH 7.3, 310 mOsm/kg) and an intracellular (pipette) solution composed of 150 mM KCl, 10 mM NaCl, 1 mM EDTA-Na, 10 mM HEPES-Na/HCl (pH 7.45, 310 mOsm/kg). Single-channel recordings were done in excised outside-out mode. Stock eugenol solution was prepared by dissolving in DMSO at 150 mM, and working solutions were prepared by diluting stocks to 3 μM in extracellular solution. Solutions were locally perfused using a microperfusion system (ALA Scientific Instruments).

Cell-based GCaMP fluorescence calcium flux assay

All DNA constructs used in this assay were cloned into a modified pME18 s vector with no fluorescent marker, flanked by AscI/NotI restriction enzyme sites for efficient cloning. Each transfection condition contained 0.5 μg of a plasmid encoding GCaMP6s (Addgene #40753) and 1.5 μg of the plasmid encoding the appropriate

olfactory receptor, diluted in 250 µl OptiMEM (Gibco). In experiments with heteromeric olfactory receptors, the total amount of DNA was 1.5 µg, in a ratio of 1:1 of Orco:OR. These were diluted in a solution containing 7 µl Lipofectamine 2000 (Invitrogen) and 250 µl OptiMem, followed by a 20-min incubation at room temperature. HEK293 cells were maintained in high-glucose DMEM supplemented with 10% (v/v) FBS and 1% (v/v) GlutaMAX at 37 °C with 5% (v/v) carbon dioxide. Cells were detached using trypsin and resuspended to a final concentration of 1×10^6 cells/ml. Cells were added to each transfection condition, mixed and added to 2×16 wells in a 384-well plate (Grenier CELLSTAR). Four to six hours later, a 16-port vacuum manifold on low vacuum was used to remove the transfection medium, replaced by fresh FluoroBrite DMEM (Gibco) supplemented with 10% (v/v) FBS and 1% (v/v) GlutaMAX. Twenty-four hours later, this medium was replaced with 20 µl reading buffer (20 mM HEPES/NaOH (pH 7.4), 1× HBSS (Gibco), 3 mM Na₂CO₃, 1 mM MgSO₄, and 2 or 5 mM CaCl₂) in each well. The calcium concentration was optimized for each receptor to account for their differences in baseline activity: for experiments with *MhOR5* and *MhOR5* mutants, reading buffer contained 2 mM CaCl₂, while 5 mM CaCl₂ was used for *MhOR1*, Orco and Orco–AgOR28 heteromers. The fluorescence emission at 527 nm, with excitation at 480 nm, was continuously read by a Hamamatsu FDSS plate reader. After 30 s of baseline recording, an optimized amount of odorant solution—10 µl for all *MhOR*-containing experiments or 20 µl for all Orco-containing experiments—was added to the cells and read for 2 min. All solutions were warmed to 37 °C before beginning.

Seven ligand concentrations were used for each transfection condition in sequential dilutions of 3, alongside a control well of only reading buffer. Ligands were dissolved in DMSO to 150 mM, then diluted with reading buffer to a highest final-well concentration of 0.5 mM (DMSO never exceeded 0.5%). Water-soluble ligands (arabinose, caffeine, denatonium, glucose, MSG, sucrose) were dissolved directly into reading buffer. If experimental data indicated a more sensitive response than this range, the concentration was adjusted accordingly. Ligand concentrations for mutants were the same as for the corresponding wild-type OR. Each plate contained a negative control of GCaMP6s transfected alone and exposed to eugenol for *MhOR5* and VUAA1 for Orco experiments. Additionally, each plate included the corresponding wild-type OR with its cognate ligand—*MhOR5* and *MhOR1* with eugenol, Orco with VUAA1, and Orco–AgOR28 with acetophenone—as a positive control to account for plate-to-plate variation in transfection efficiency and cell count. A control of DMSO alone was also tested to ensure no activity effects were due to the solvent. Each concentration of ligand was applied to four technical replicates, which were averaged and considered a single biological replicate.

The baseline fluorescence (F) was calculated as the average fluorescence of the 30 s before odour was added to the plate. Within each well, ΔF was calculated as the

difference between the average of the last 10 s of fluorescence and the baseline F . $\Delta F/F$ was then calculated as the ΔF divided by the baseline fluorescence (F). Finally, the $\Delta F/F$ for each concentration was normalized to the maximum $\Delta F/F$ value of the corresponding positive control present on each plate: *MhOR5* and *MhOR1* with eugenol, Orco with VUAA1, and Orco–AgOR28 with acetophenone to account for inevitable variations in transfection efficiency and cell counts across different plates. The normalized $\Delta F/F$ averaged across all experiments for a given condition is the value used to construct the dose–response curves in all plots (Figs. 1b, 2e–g, Extended Data Figs. 2d, 9a–c, 10c, 11b). All wild-type curves come from the same plates as the experimental data in the same plot. Baseline values for wild-type and mutant channels were found by normalizing each F value by the negative GCaMP6s-only control on the same plate (Extended Data Figs. 1c, 9a,e).

For all experiments, GraphPad Prism 8 was used to fit the dose–responses curves to the Hill equation from which the EC_{50} of the curve was extracted. Three metrics were used to characterize the dose–response curve for each ligand: activity index, $\log(EC_{50})$ and max $\Delta F/F$. For conditions where EC_{50} was too high for the dose–response curve to reach saturation and therefore could not be fitted to a Hill equation, a value of –2 was assigned to the EC_{50} , which is more than an order of magnitude higher than the highest concentration used. Max $\Delta F/F$ is the maximum response achieved at the highest concentration. Activity index is defined as the negative product of $\log(EC_{50})$ and max $\Delta F/F$, as follows:

$$\text{Activity index} = -\log(EC_{50}) \times \max \Delta F/F$$

Gels and small-scale transfections

For western blots and fluorescence-detection size-exclusion chromatography (FSEC) traces (Extended Data Figs. 1a, b, 9g), HEK293 cells were maintained in high-glucose DMEM supplemented with 10% (v/v) FBS and 1% (v/v) GlutaMAX at 37 °C with 5% (v/v) carbon dioxide. Cells were detached using trypsin and plated in six-well plates at a concentration of 0.4×10^6 per well. Twenty-four hours later, cells were transfected with 2 µg of DNA in the same superfolder GFP-containing pEG BacMam vector used for large-scale purification and 9 µl Lipofectamine 2000 (Invitrogen) diluted in 700 µl OptiMEM and added dropwise to the cells after a 5-min incubation. Twenty-four hours later, cells were checked for GFP fluorescence, rinsed with phosphate-buffered saline, and collected by centrifugation. Cells were either frozen at –20 °C or used immediately.

Cell pellets were rapidly thawed and resuspended in 200 µl lysis buffer containing 50 mM HEPES/NaOH (pH 7.5), 375 mM NaCl, an EDTA-free protease inhibitor cocktail

(Roche), and 1 mM PMSF. The protein was extracted for 2 h at 4 °C by adding 0.5% (w/v) DDM with 0.1% (w/v) CHS after 10 s sonication in a water bath. This mixture was then clarified by centrifugation and filtered. The supernatant was added to a Shimadzu autosampler connected to a Superose 6 Increase column equilibrated in SEC buffer. An aliquot of the supernatant was also used to run SDS–PAGE (Bio-Rad, 12% Mini-PROTEAN TGX) and Blue Native(BN)-PAGE (Invitrogen, 3–12% Bis-Tris) gels. Gels were transferred using Trans-Blot Turbo Transfer Pack (Bio-Rad) and blocked overnight. The following day, gels were stained with rabbit anti-GFP polyclonal antibody (Life Technologies; 1:20,000), washed, incubated with anti-rabbit secondary antibody (1:10,000), and imaged with ImageLab.

Lifetime sparseness calculation

The lifetime sparseness^{54,55} measure in Extended Data Fig. 1d was used to quantify olfactory receptor tuning breadth and calculated as follows:

$$\text{Lifetime sparseness} = \left(\frac{1}{n} - \frac{1}{n} \right) \times \left(1 - \frac{\sum_{i=1}^n \frac{\text{res}_i}{\text{res}_i}}{\left(\sum_{i=1}^n \frac{\text{res}_i}{\text{res}_i} \right)^2} \right)$$

in which n is the number of ligands in the set, and res_i is the receptor's response to a given ligand i . All inhibitory responses (values below 0) were set to 0 before the calculation^{54,55}. The *Drosophila melanogaster* OR dataset comes from the DoOR database⁵⁶.

Multiple regression analysis

A set of 11 molecular descriptors were compiled for all 54 ligands tested from PubChem, Sigma-Aldrich, ChemSpider, EPA, and The Good Scents Company; the values used are in Supplementary Table 9. A multiple regression analysis using the scikit-learn Linear Regression module was used to assess the accuracy with which the receptor activity could be predicted by individual descriptors (1-dimensional analysis) or combinations of two descriptors (2-dimensional analysis) (Extended Data Table 2). Owing to the absence of reported metrics for some ligands—acetic acid, citric acid, MSG, sucrose, denatonium, and VUAA1—the analysis was performed on the remaining 48 ligands. For the 1-dimensional analysis, a single variable linear regression was performed for each descriptor independently. The analysis sought to fit a linear model with coefficients $w_1, \dots, w_n + 1$, in which n is the dimension of the input data. The optimal coefficient set was determined using residual sum of squares optimization between the observed activity index targets and those predicted by linear approximation using solved coefficients. This process was repeated for the 2-dimensional case, using every unique permutation of descriptors across the 11-

dimensional space. As a means of assessing the predictive power of a given combination, the R^2 -value, reflecting the square of the correlation coefficient between observed and modelled values of the activity index, was calculated for each linear model and reported in Extended Data Table 2. This allowed ranking of descriptor sets based on accuracy of prediction.

Sequence alignments

For Extended Data Fig. 11a, the alignment between the sequences of *MhOR1* and *MhOR5* was done using MAFFT implemented in JalView⁵⁷ with minimal manual adjustment based on the structure of *MhOR5*. For Extended Data Fig. 5a, the sequence alignment between *A. bakeri* Orco and *MhOR5* was done by aligning the published structure of *A. bakeri* Orco (PDB 6C70) and the structure of *MhOR5* in PyMOL. All sequence alignments were visualized and plotted using JalView⁵⁷.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All data underlying this study are available upon request from the corresponding author.

References

1. 1.

Bargmann, C. I. Comparative chemosensation from receptors to ecology. *Nature* **444**, 295–301 (2006).

2. 2.

Robertson, H. M. Molecular evolution of the major arthropod chemoreceptor gene families. *Annu. Rev. Entomol.* **64**, 227–242 (2019).

3. 3.

Bear, D. M., Lassance, J. M., Hoekstra, H. E. & Datta, S. R. The evolving neural and genetic architecture of vertebrate olfaction. *Curr. Biol.* **26**, R1039–R1049

(2016).

4. 4.

Brand, P. et al. The origin of the odorant receptor gene family in insects. *eLife* **7**, e38340 (2018).

5. 5.

Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).

6. 6.

Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5**, 124–133 (2002).

7. 7.

Sato, K. et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**, 1002–1006 (2008).

8. 8.

Wicher, D. et al. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452**, 1007–1011 (2008).

9. 9.

Larsson, M. C. et al. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).

10. 10.

Butterwick, J. A. et al. Cryo-EM structure of the insect olfactory receptor Orco. *Nature* **560**, 447–452 (2018).

11. 11.

Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).

12. 12.

Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).

13. 13.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. Odor coding by a Mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).

14. 14.

Thoma, M. et al. Transcriptome surveys in silverfish suggest a multistep origin of the insect odorant receptor gene family. *Front. Ecol. Evol.* **7**, 1–13 (2019).

15. 15.

Jones, P. L., Pask, G. M., Rinker, D. C. & Zwiebel, L. J. Functional agonism of insect odorant receptor ion channels. *Proc. Natl Acad. Sci. USA* **108**, 8821–8825 (2011).

16. 16.

Kenakin, T. A scale of agonism and allosteric modulation for assessment of selectivity, bias, and receptor mutation. *Mol. Pharmacol.* **92**, 414–424 (2017).

17. 17.

Wang, G., Carey, A. F., Carlson, J. R. & Zwiebel, L. J. Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **107**, 4418–4423 (2010).

18. 18.

Robertson, H. M. The insect chemoreceptor superfamily is ancient in animals. *Chem. Senses* **40**, 609–614 (2015).

19. 19.

Pask, G. M., Jones, P. L., Rützler, M., Rinker, D. C. & Zwiebel, L. J. Heteromeric Anopheline odorant receptors exhibit distinct channel properties. *PLoS ONE* **6**, e28774 (2011).

20. 20.

Halgren, T. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759 (2004).

21. 21.

Vincent, F. et al. Complexes of porcine odorant binding protein with odorant molecules belonging to different chemical classes. *J. Mol. Biol.* **300**, 127–139 (2000).

22. 22.

Brito, N. F., Moreira, M. F. & Melo, A. C. A. A look inside odorant-binding proteins in insect chemoreception. *J. Insect Physiol.* **95**, 51–65 (2016).

23. 23.

Hopf, T. A. et al. Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. *Nat. Commun.* **6**, 6077 (2015).

24. 24.

Yuan, S. et al. Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun. Biol.* **2**, 141 (2019).

25. 25.

Araneda, R. C., Kini, A. D. & Firestein, S. The molecular receptive range of an odorant receptor. *Nat. Neurosci.* **3**, 1248–1255 (2000).

26. 26.

Amoore, J. E. Stereochemical theory of olfaction. *Nature* **198**, 271–272 (1963).

27. 27.

Auer, T. O. et al. Olfactory receptor and circuit evolution promote host specialization. *Nature* **579**, 402–408 (2020).

28. 28.

Shaw, K. H., Johnson, T. K., Anderson, A., de Bruyne, M. & Warr, C. G. Molecular and functional evolution at the odorant receptor *Or22* locus in

Drosophila melanogaster. *Mol. Biol. Evol.* **36**, 919–929 (2019).

29. 29.

Yang, K., Huang, L. Q., Ning, C. & Wang, C. Z. Two single-point mutations shift the ligand selectivity of a pheromone receptor between two closely related moth species. *eLife* **6**, 1–21 (2017).

30. 30.

Yuvaraj, J. K. et al. Putative ligand binding sites of two functionally characterized bark beetle odorant receptors. *BMC Biol.* **19**, 16 (2021).

31. 31.

DeGennaro, M. The mysterious multi-modal repellency of DEET. *Fly (Austin)* **9**, 45–51 (2015).

32. 32.

Wilson, R. I. & Mainen, Z. F. Early events in olfactory processing. *Annu. Rev. Neurosci.* **29**, 163–201 (2006).

33. 33.

Charlier, L. et al. How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell. Mol. Life Sci.* **69**, 4205–4213 (2012).

34. 34.

Baud, O. et al. Exchanging ligand-binding specificity between a pair of mouse olfactory receptor paralogs reveals odorant recognition principles. *Sci. Rep.* **5**, 14948 (2015).

35. 35.

Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J. Neurosci.* **25**, 1806–1815 (2005).

36. 36.

Bohbot, J. D. & Dickens, J. C. Selectivity of odorant receptors in insects. *Front. Cell. Neurosci.* **6**, 29 (2012).

37. 37.

Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protocols* **9**, 2574–2585 (2014).

38. 38.

Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).

39. 39.

Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

40. 40.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

41. 41.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

42. 42.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

43. 43.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

44. 44.

Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J.*

Mol. Biol. **333**, 721–745 (2003).

45. 45.

Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.* **54**, 5.6.1–5.6.37 (2016).

46. 46.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

47. 47.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

48. 48.

Paknejad, N. & Hite, R. K. Structural basis for the regulation of inositol trisphosphate receptors by Ca^{2+} and IP_3 . *Nat. Struct. Mol. Biol.* **25**, 660–668 (2018).

49. 49.

Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

50. 50.

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

51. 51.

Release 2020-4: Glide, Schrödinger, LLC (2020).

52. 52.

Shelley, J. C. et al. Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21**, 681–691 (2007).

53. 53.

Smart, O. S., Neduvvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, 376 (1996).

54. 54.

Grabe, V. et al. Elucidating the neuronal architecture of olfactory glomeruli in the *Drosophila* antennal lobe. *Cell Rep.* **16**, 3401–3413 (2016).

55. 55.

Schlief, M. L. & Wilson, R. I. Olfactory processing and behavior downstream from highly selective receptor neurons. *Nat. Neurosci.* **10**, 623–630 (2007).

56. 56.

Münch, D. & Galizia, C. G. DoOR 2.0—comprehensive mapping of *Drosophila melanogaster* odorant responses. *Sci. Rep.* **6**, 21841 (2016).

57. 57.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

58. 58.

Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).

Acknowledgements

We thank R. Axel, R. MacKinnon, J. Chen, S. Klinge, B. Noro, J. Butterwick, M. Maldonado, C. McBride and members of the Ruta laboratory for discussion and comments on the manuscript, R. Hite and T. Walz for advice on cryo-EM processing; P. Stock for advice on the multiple regression analysis; P. Brand for advice on receptor alignments; L. Ramos-Espiritu for support on functional assays; and M. Ebrahim and J. Sotiris at The Rockefeller University Evelyn Gruss Lipper Cryo-Electron Microscopy Resource Center for assistance with microscope operation. This work was supported by a Leon Levy Postdoctoral Fellowship (to J.d.M.) and the National Institutes of Health (K99DC019401 to J.d.M. and R01AI103171 to V.R.).

Author information

Affiliations

1. Laboratory of Neurophysiology and Behavior, The Rockefeller University, New York, NY, USA

Josefina del Mármol, Mackenzie A. Yedlin & Vanessa Ruta

Contributions

V.R. and J.d.M. conceived the study, designed experiments and wrote the manuscript, with input from M.A.Y. J.d.M. purified *MhOR5*, performed electrophysiology, collected and analysed cryo-EM data, built and refined the models, and performed molecular docking. M.A.Y. performed molecular biology, cell culture and calcium imaging assays.

Corresponding author

Correspondence to [Vanessa Ruta](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Rachelle Gaudet, Emily Liman 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 Biochemical and functional comparison of *MhOR5* and *MhOR1*.**](#)

a, Representative image of western blots of all five *Mh*ORs and *A. bakeri* Orco fused to GFP in non-denaturing Blue Native gels, stained with anti-GFP antibodies. Position of the Orco tetramer, dimer and monomer marked to indicate stoichiometry. Full lanes are shown. This experiment was repeated five times with comparable results. **b**, Size-exclusion chromatography (SEC) trace of *Mh*OR5. **c**, Baseline fluorescence of *Mh*OR5 and *Mh*OR1 in the functional calcium assay, normalized to a GCaMP-only control on the same plate ($n = 12$). Statistical significance was determined using an unpaired, two-tailed *t*-test. $****P < 0.0001$. **d**, Comparison of breadth of tuning, measured as lifetime sparseness (see [Methods](#)), of *D. melanogaster* ORs and *Mh*ORs. Lifetime sparseness close to 0 suggests broad tuning whereas 1 indicates selective tuning to a single ligand in the panel. **e**, **f**, Receptor activity for *Mh*OR5 (**e**, top) and *Mh*OR1 (**e**, bottom) with ligands (named in **f**) sorted by chemical classes. n values for biological replicates provided in Supplementary Tables [1](#), [6](#), [7](#). Data are presented as mean values \pm s.e.m.

Extended Data Fig. 2 *Mh*OR5 is activated by a broad set of odorants.

a–c, Tuning curves of *Mh*OR5 sorted by activity index (**a**), max $\Delta F/F$ (**b**), and $-\log(\text{EC}_{50})$ (**c**). Activity index was calculated as $-\log(\text{EC}_{50}) \times \text{max } \Delta F/F$. In **c**, the inhibitory ligands VUAA1 and denatonium are shown as IC_{50} . More details about these measurements in Methods. **d**, Dose–response curves for all the individual ligands shown in **a–c**, averaged across all experiments and shown with s.e.m. Note that the activity index, which combines $-\log(\text{EC}_{50})$ and max $\Delta F/F$, captures agonism by ligands with low affinity such as 1-pentanol (teal) or sub-maximal efficacy such as DEET (green). Relevant summary tables, including n values for biological replicates, are in Supplementary Tables [1](#), [8](#). **e**, Representative traces of an excised outside-out patch of HEK cells expressing *Mh*OR5 and stimulated with 3 μM eugenol at voltages from -60 mV to 40 mV. Right, current–voltage (*IV*) plot of eugenol-elicited currents from whole-cell recordings of HEK cells expressing *Mh*OR5 (shown as absolute current measurements without normalization, with s.d. from n independent replicates; $n_{-80 \text{ mV}} = 8$, $n_{-40 \text{ mV}} = 7$, $N_0 \text{ mV} = 7$, $n_{40 \text{ mV}} = 4$).

Extended Data Fig. 3 Cryo-EM data analysis.

a–c, Processing pipelines for the apo structure (**a**), the eugenol-bound structure (**b**), and the DEET-bound structure (**c**). Details can be found in Extended Data Table [1](#).

Extended Data Fig. 4 Cryo-EM density with map and model analysis and validation.

a–c, Left, cryo-EM densities for the modelled regions of the eugenol-bound structure (**a**), the DEET-bound structure (**b**), and the apo structure (**c**). Models are shown in stick representation within the density, with the helices denoted underneath from the N terminus (S0) to the C terminus (S7b). Right, local resolution estimation for each final map, calculated in cryoSPARC v2. Side views shown with front and back subunits removed for visualization. **d–f**, Top, FSC curves for the final cryo-EM density maps, obtained with cryoSPARC v2. The horizontal dashed line intersects at 0.143, the cutoff value. Bottom, FSC relationships between final map and model (black, sum), half-map 1 and model (red, work), and half-map 2 and model (blue, free), calculated in phenix.mtriage.

[Extended Data Fig. 5 Conserved architecture of insect olfactory receptors.](#)

a, Sequence alignment of *MhOR5* and *A. bakeri* Orco. Sequence identity (dark purple) and similarity (light purple) are highlighted. The positions of the helices in *MhOR5* are marked, as well as the S4–S5 intracellular loop. **b**, Structural overlay of *AbOrco* (gold) and the apo state of *MhOR5* (blue) from the side view (left) with grey bars indicating the positions at which cross-sections are taken for insets shown from top views (top right) and anchor views (bottom right). The r.m.s.d. between individual subunits is 3.00 Å, and superposition of the whole tetramers yields an r.m.s.d. of 3.18 Å (calculated using PDBeFold server). **c**, Dose–response curves of wild-type and mutant *MhOR5* (left) and *AbOrco* (right), in which the S4–S5 intracellular loop (unmodelled in both *MhOR5* and *AbOrco* structures) has been replaced by a short linker (Δ S4–S5 loop), in response to eugenol ($n = 6$) and VUAA1 ($n = 4$), respectively. Error bars denote s.e.m. **d**, Inter-subunit interactions are concentrated in the anchor region. Top, side view of *MhOR5* with front and back subunits removed for visualization. Residues within 5 Å of residues in neighbouring subunits are shows as spheres, coloured by subunit. Insets (below) show extracellular views of cross-sections taken at the top and anchor positions as indicated by grey bars in the side view.

[Extended Data Fig. 6 Detail of the odorant-binding region in *MhOR5*.](#)

a, The binding pocket and ligand are shown in three orientations. The position of the pocket in each panel view is shown embedded within the tetrameric receptor to aid with orientation. Residues are colour-coded as in the bar graph of Fig. 3c, displaying the position of the ‘control’ residues L379 and T87 in purple that project away from the binding site and the binding site residues in yellow. The eugenol molecule is shown in grey. **b**, Schematic of the position of the three different views of the binding pocket. Model shown as ribbon and density shown as black mesh of the odorant-binding region of the apo (top), eugenol-bound (middle) and DEET-bound (bottom) structures. Cryo-EM density contoured at same level in all panels.

Extended Data Fig. 7 Electrostatic surface representation of the eugenol-bound model.

a, Top and side views of *MhOR5* bound to eugenol, with detailed views of the ligand binding cavity. **b**, Cross-sections of eugenol-bound *MhOR5* at the positions indicated with planes, highlighting the electrostatic environment of the side exits and the occlusion at the level of the anchor domain. Potentials estimated in ChimeraX v1.1 using coulombic calculation.

Extended Data Fig. 8 Docking of *MhOR5* agonists in the binding pocket.

a, View of the binding pocket in the eugenol-bound (top, pink) and DEET-bound (middle, teal) *MhOR5* structures with the representative top poses for eugenol and DEET. Cryo-EM densities for eugenol and DEET shown as black mesh. The docking scores (calculated using Glide) of the eugenol poses (top) are, from left to right, -6,59, -6,71 and -6,81; and of the DEET poses (middle) are -7,33 (left) and -7,37 (right). Bottom, the top poses for each of the agonists of *MhOR5* from Glide are presented overlaid according to chemical class. **b**, Chemical structures of all activators in **a**.

Extended Data Fig. 9 Effects of binding pocket mutations on *MhOR5* function.

a, Top, dose-response curves of *MhOR5* mutants in S2, S3 and S6 helices and two residues adjacent to the binding pocket but projecting away from it (denoted as control). Each WT curve represents the corresponding controls from the same experiments. All dose-response curves represent the average across experiments, error bars denote s.e.m. Bottom, baseline fluorescence of each mutant normalized to GCaMP-only control on the same plate. For all baseline data, columns are presented as mean values \pm s.e.m. and statistical significance was determined using one-way ANOVA followed by Dunnett's multiple comparison tests comparing mutants to the respective wild-type *MhOR5* control experiments. For mutants where the EC₅₀ was incalculably high and Bartlett's test showed non-homogenous variance, statistical significance was determined with a Brown-Forsythe test. **b, c**, Dose-response curves for WT, M209 mutant series and I213 mutant series of *MhOR5* in response to eugenol (pink) and DEET (teal). **d**, Detailed views of the binding pocket in the eugenol-bound (left, pink) and DEET-bound (right, teal) *MhOR5* structures. **e**, For each mutant shown in **b, c**, baseline fluorescence of each mutant normalized to GCaMP-only control on the same plate. **f**, Tuning curves of M209V, I213M and WT *MhOR5* showing $-\log(\text{EC}_{50})$, ordered as in Fig. 4e. **g**, Western blots of denaturing and native gels showing protein expression and multimeric assembly for each mutant, stained with anti-GFP. Full lanes shown and all gels run in parallel from samples in the same

experiment. For all baseline measurements, baseline represents the mean of the first 30 s of fluorescence before the addition of any ligand. Relevant summary tables, including n values for biological replicates, are in Supplementary Tables 2, 4.
**** $P < 0.0001$; *** $P < 0.0005$; ** $P < 0.005$, * $P < 0.05$, NS, not significant.

Extended Data Fig. 10 A potential route for coupling of odorant binding to pore opening and details of pore opening in the DEET-bound *MhOR5* structure.

a, Top view of *MhOR5* with helices represented as tubes in apo structure (blue) and eugenol-bound structure (pink). **b**, Close-up view of one subunit of the apo (blue, top) and eugenol-bound (pink, bottom) structures of *MhOR5*. Residues Leu465 (in S7b), Tyr362 (in S5) and Tyr380 (in S6, lining the binding pocket) and eugenol are shown as sticks with a translucent outline of the sphere representation. **c**, Mutation of Leu465 in S7 and Tyr362 in S5 into alanine impairs receptor function. A conservative substitution of Tyr362 to phenylalanine restores wild-type activity, highlighting the potential role of hydrophobic packing in connecting odorant binding with pore opening ($n = 6$). Dose–response curves represent the average across experiments, error bars denote s.e.m. Summary table of receptor data is in Supplementary Table 2. **d**, The lumen of the central pathway and side exits of the DEET-bound structure of *MhOR5*. **e**, Diameter of the ion pathway beginning at the extracellular membrane (position 0) and following the ion conduction pathway (solid line) and the central four-fold axis through the anchor domain (dashed line), calculated with HOLE. Blue line represents apo structure, pink line represents eugenol-bound structure and cyan line represents DEET-bound structure. **f**, Top view of the DEET-bound structure with inset (right) highlighting the positions of residues Gln467 and Val468.

Extended Data Fig. 11 Putative binding pocket in *MhOR1*.

a, Sequence alignment of *MhOR5* and *MhOR1*, highlighting the position of the S0–S7 helices in *MhOR5* and the location of the binding pocket residues in *MhOR5* and homologous residues in *MhOR1*, colour-coded as in **b**. **b**, Dose–response curves of *MhOR1* mutants ($n = 3$). Curves for mutants T103A, Y106A and I107A with eugenol, G169A and W173A with eugenol, and L227A and M231A with eugenol overlap almost completely and therefore appear overlaid. Dose–response curves represent the average across experiments, error bars denote s.e.m. **c**, Mutations of putative binding pocket residues in *MhOR1* significantly and differentially alter ligand binding for the two best *MhOR1* ligands, 1-octanol and eugenol. Mutation of the putative ‘control’ residues in *MhOR1* (homologous to those that in *MhOR5* are adjacent to the pocket but projecting away from the ligand) are also not directly involved in ligand binding in *MhOR1* ($n = 3$). Statistical significance determined using one-way ANOVAs followed by Dunnett’s multiple comparison tests, comparing mutants to their respective wild-

type controls for each ligand. For mutants where the EC₅₀ was incalculably high and Bartlett's test showed non-homogenous variance, statistical significance was determined with a Brown–Forsythe test. ****P < 0.0001; ***P < 0.001; NS, not significant. Data shown as mean ± s.e.m. **d**, EC₅₀ and activity index tuning curves for wild-type *MhOR1* and a conservative binding pocket mutant, M231I, against 17 diverse ligands. Ligands are ordered high-to-low by EC₅₀ and activity index scores for the wild-type *MhOR1*. Data shown as mean ± s.e.m. Relevant summary tables, including *n* values for the biological replicates, can be found in Supplementary Tables [10](#), [11](#).

Extended Data Table 1 Cryo-EM data collection, refinement and model statistics

Extended Data Table 2 The correlation between chemical metrics and receptor activity for *MhOR5* and *MhOR1*

Supplementary information

Supplementary Information

This file contains supplementary text, supplementary figure 1 and supplementary tables 1 – 11.

Reporting Summary

Supplementary Video 1

Morph video of MhOR5 showing both apo and bound states from the top view.

Supplementary Video 2

Morph video of MhOR5 showing both apo and bound states from the side view.

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

del Mármol, J., Yedlin, M.A. & Ruta, V. The structural basis of odorant recognition in insect olfactory receptors. *Nature* **597**, 126–131 (2021).
<https://doi.org/10.1038/s41586-021-03794-8>

- Received: 28 January 2021
- Accepted: 01 July 2021
- Published: 04 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03794-8>

[Insights into a receptor that lets insects sense scents](#)

- Emily R. Liman

News & Views 04 Aug 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03794-8>

- Article
- [Published: 18 August 2021](#)

DAXX represents a new type of protein-folding enabler

- [Liangqian Huang](#) [ORCID: orcid.org/0000-0002-7146-0296^{1,2}](#),
- [Trisha Agrawal^{1,2}](#) [nAff5](#),
- [Guixin Zhu^{1,2}](#),
- [Sixiang Yu^{1,2}](#),
- [Liming Tao³](#),
- [JiaBei Lin](#) [ORCID: orcid.org/0000-0001-8863-1859⁴](#),
- [Ronen Marmorstein](#) [ORCID: orcid.org/0000-0003-4373-4752^{2,4}](#),
- [James Shorter⁴](#) &
- [Xiaolu Yang](#) [ORCID: orcid.org/0000-0001-6755-7725^{1,2}](#)

[Nature](#) volume 597, pages 132–137 (2021)

- 10k Accesses
- 98 Altmetric
- [Metrics details](#)

Subjects

- [Cancer](#)
- [Neurological disorders](#)
- [Protein folding](#)
- [Tumour-suppressor proteins](#)

Abstract

Protein quality control systems are crucial for cellular function and organismal health. At present, most known protein quality control systems are multicomponent machineries that operate via ATP-regulated interactions with non-native proteins to prevent aggregation and promote folding¹, and few systems that can broadly enable protein folding by a different mechanism have been identified. Moreover, proteins that contain the extensively charged poly-Asp/Glu (polyD/E) region are common in eukaryotic proteomes², but their biochemical activities remain undefined. Here we show that DAXX, a polyD/E protein that has been implicated in diverse cellular processes^{3,4,5,6,7,8,9,10}, possesses several protein-folding activities. DAXX prevents aggregation, solubilizes pre-existing aggregates and unfolds misfolded species of model substrates and neurodegeneration-associated proteins. Notably, DAXX effectively prevents and reverses aggregation of its in vivo-validated client proteins, the tumour suppressor p53 and its principal antagonist MDM2. DAXX can also restore native conformation and function to tumour-associated, aggregation-prone p53 mutants, reducing their oncogenic properties. These DAXX activities are ATP-independent and instead rely on the polyD/E region. Other polyD/E proteins, including ANP32A and SET, can also function as stand-alone, ATP-independent molecular chaperones, disaggregases and unfoldases. Thus, polyD/E proteins probably constitute a multifunctional protein quality control system that operates via a distinctive mechanism.

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: DAXX prevents protein misfolding and aggregation.

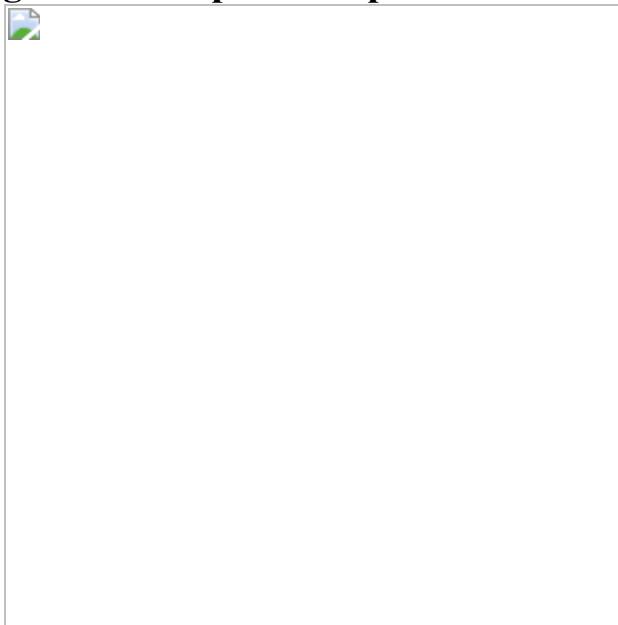


Fig. 2: DAXX dissolves protein aggregates and unfolds misfolded species.

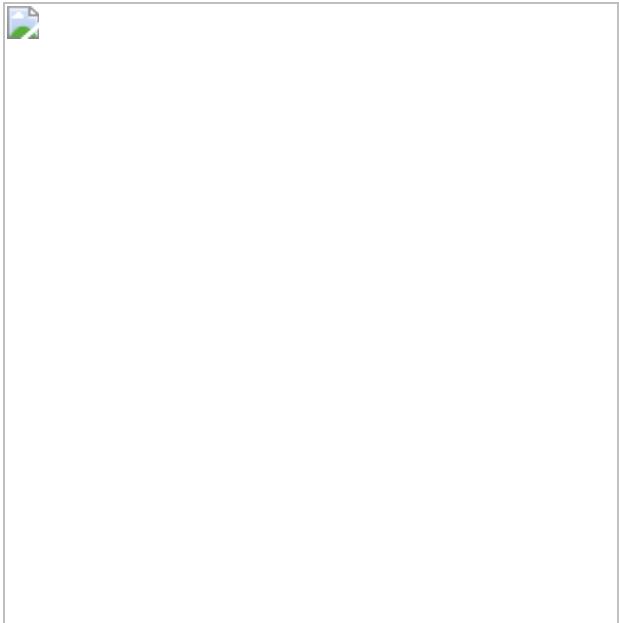


Fig. 3: Other polyD/E proteins can function as molecular chaperones, disaggregases, and unfoldases.

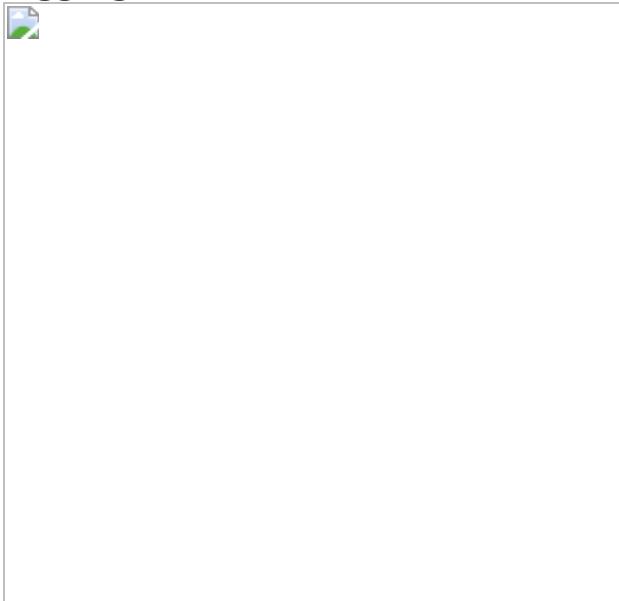
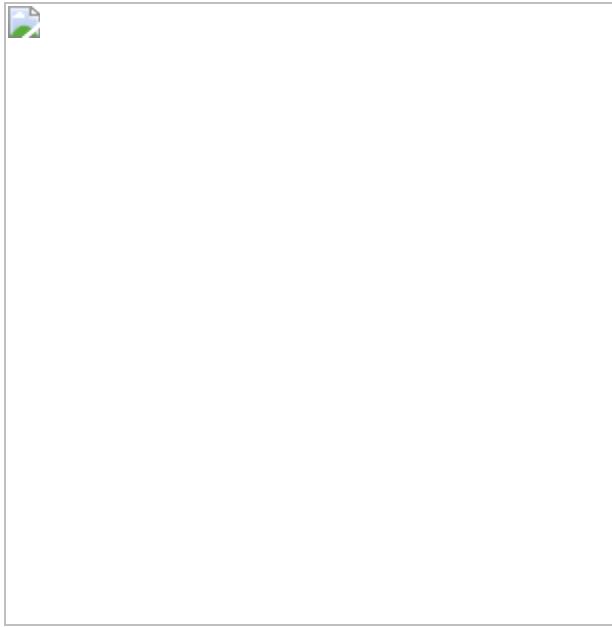


Fig. 4: DAXX maintains and restores the native conformation of p53 and MDM2.



Data availability

All data supporting the findings of this study are provided within the manuscript and its [Supplementary Information](#). [Source data](#) are provided with this paper.

Code availability

Source code and datasets for polyD/E protein analysis are available in GitHub: <https://github.com/SunmoonTao/de-enrichemnt-analysis>.

References

1. 1.

Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).

2. 2.

Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J. & Gentles, A. J. Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl Acad. Sci. USA* **99**, 333–338 (2002).

3. 3.

Yang, X., Khosravi-Far, R., Chang, H. Y. & Baltimore, D. Daxx, a novel Fas-binding protein that activates JNK and apoptosis. *Cell* **89**, 1067–1076 (1997).

4. 4.

Chang, H. Y., Nishitoh, H., Yang, X., Ichijo, H. & Baltimore, D. Activation of apoptosis signal-regulating kinase 1 (ASK1) by the adapter protein Daxx. *Science* **281**, 1860–1863 (1998).

5. 5.

Perlman, R., Schiemann, W. P., Brooks, M. W., Lodish, H. F. & Weinberg, R. A. TGF-beta-induced apoptosis is mediated by the adapter protein Daxx that facilitates JNK activation. *Nat. Cell Biol.* **3**, 708–714 (2001).

6. 6.

Zhao, L. Y. et al. Negative regulation of p53 functions by Daxx and the involvement of MDM2. *J. Biol. Chem.* **279**, 50566–50579 (2004).

7. 7.

Tang, J. et al. Critical role for Daxx in regulating Mdm2. *Nat. Cell Biol.* **8**, 855–862 (2006).

8. 8.

Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl Acad. Sci. USA* **107**, 14075–14080 (2010).

9. 9.

Song, M. S. et al. The deubiquitylation and localization of PTEN are regulated by a HAUSP-PML network. *Nature* **455**, 813–817 (2008).

10. 10.

Mahmud, I. & Liao, D. DAXX in cancer: phenomena, processes, mechanisms and regulation. *Nucleic Acids Res.* **47**, 7734–7752 (2019).

11. 11.

Michaelson, J. S., Bader, D., Kuo, F., Kozak, C. & Leder, P. Loss of Daxx, a promiscuously interacting protein, results in extensive apoptosis in early mouse development. *Genes Dev.* **13**, 1918–1923 (1999).

12. 12.

Jiao, Y. et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**, 1199–1203 (2011).

13. 13.

Gopal, R. K. et al. Widespread chromosomal losses and mitochondrial DNA alterations as genetic drivers in hurthle cell carcinoma. *Cancer Cell* **34**, 242–255 (2018).

14. 14.

Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).

15. 15.

Jackrel, M. E. et al. Potentiated Hsp104 variants antagonize diverse proteotoxic misfolding events. *Cell* **156**, 170–182 (2014).

16. 16.

Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45–51 (2013).

17. 17.

Kayed, R. et al. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–489 (2003).

18. 18.

Sabil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630–642 (2013).

19. 19.

Glover, J. R. & Lindquist, S. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* **94**, 73–82 (1998).

20. 20.

Sharma, S. K., De los Rios, P., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **6**, 914–920 (2010).

21. 21.

Gupta, R. et al. Firefly luciferase mutants as sensors of proteome stress. *Nat. Methods* **8**, 879–884 (2011).

22. 22.

Outeiro, T. F. et al. Formation of toxic oligomeric alpha-synuclein species in living cells. *PLoS ONE* **3**, e1867 (2008).

23. 23.

Rentzepis, D., Jonsson, T. & Sauer, R. T. Acceleration of the refolding of Arc repressor by nucleic acids and other polyanions. *Nat. Struct. Biol.* **6**, 569–573 (1999).

24. 24.

Walker, J. M., Hastings, J. R. & Johns, E. W. A novel continuous sequence of 41 aspartic and glutamic residues in a non-histone chromosomal protein. *Nature* **271**, 281–282 (1978).

25. 25.

Adachi, Y., Pavlakis, G. N. & Copeland, T. D. Identification and characterization of SET, a nuclear phosphoprotein encoded by the translocation break point in acute undifferentiated leukemia. *J. Biol. Chem.* **269**, 2258–2262 (1994).

26. 26.

Vaesen, M. et al. Purification and characterization of two putative HLA class II associated proteins: PHAPI and PHAPII. *Biol. Chem. Hoppe Seyler* **375**, 113–126 (1994).

27. 27.

Wang, D. et al. Acetylation-regulated interaction between p53 and SET reveals a widespread regulatory mode. *Nature* **538**, 118–122 (2016).

28. 28.

Butler, J. S. & Loh, S. N. Folding and misfolding mechanisms of the p53 DNA binding domain at physiological temperature. *Protein Sci.* **15**, 2457–2465 (2006).

29. 29.

Ishimaru, D. et al. Fibrillar aggregates of the tumor suppressor p53 core domain. *Biochemistry* **42**, 9022–9027 (2003).

30. 30.

Ano Bom, A. P. et al. Mutant p53 aggregates into prion-like amyloid oligomers and fibrils: implications for cancer. *J. Biol. Chem.* **287**, 28152–28162 (2012).

31. 31.

Zhang, R. & Monsma, F. Fluorescence-based thermal shift assays. *Curr. Opin. Drug Discov. Devel.* **13**, 389–402 (2010).

32. 32.

Muller, P. A. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304–317 (2014).

33. 33.

Guo, L. et al. A cellular system that degrades misfolded proteins and protects against neurodegeneration. *Mol. Cell* **55**, 15–30 (2014).

34. 34.

Zhu, G. et al. TRIM11 prevents and reverses protein aggregation and rescues a mouse model of Parkinson’s disease. *Cell Rep.* **33**, 108418 (2020).

35. 35.

Bykov, V. J. N., Eriksson, S. E., Bianchi, J. & Wiman, K. G. Targeting mutant p53 for efficient cancer therapy. *Nat. Rev. Cancer* **18**, 89–102 (2018).

36. 36.

Chu, Y. & Yang, X. SUMO E3 ligase activity of TRIM proteins. *Oncogene* **30**, 1108–1116 (2011).

37. 37.

Chen, L. et al. Enhanced degradation of misfolded proteins promotes tumorigenesis. *Cell Rep.* **18**, 3143–3154 (2017).

38. 38.

Yen, C. F., Harischandra, D. S., Kanthasamy, A. & Sivasankar, S. Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity. *Sci. Adv.* **2**, e1600014 (2016).

39. 39.

Måansson, C. et al. Interaction of the molecular chaperone DNAJB6 with growing amyloid-beta 42 (A β 42) aggregates leads to sub-stoichiometric inhibition of amyloid formation. *J. Biol. Chem.* **289**, 31066–31076 (2014).

40. 40.

Zhang, Y. et al. Upregulation of antioxidant capacity and nucleotide precursor availability suffices for oncogenic transformation. *Cell Metab.* **33**, 94–109 (2021).

Acknowledgements

We thank S. K. Sharma and P. Goloubinoff for providing the LucD plasmid, D. Brady for help with protein purification, D. Ricketts and A. Olia for help with the thermal shift assay, J. Huang for technical assistance, and E. Dean for Sf9 protein production. This work was supported by National Institutes of Health (NIH) grants R01CA182675, R01CA184867, R01CA235760, R01CA243520 (X.Y.), P01 AG031862 (R.M.) and R01GM099836 (J.S.); an Alzheimer's Association Research Fellowship and a Warren Alpert Foundation Distinguished Scholars Fellowship (J.L.); and a Sponsored Research Agreement from Wealth Strategy Holding Limited (X.Y.).

Author information

Author notes

1. Trisha Agrawal

Present address: Wilson Sonsini Goodrich & Rosati LP, New York, NY, USA

Affiliations

1. Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Liangqian Huang, Trisha Agrawal, Guixin Zhu, Sixiang Yu & Xiaolu Yang

2. Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Liangqian Huang, Trisha Agrawal, Guixin Zhu, Sixiang Yu, Ronen Marmorstein & Xiaolu Yang

3. Broad Institute of MIT and Harvard, Cambridge, MA, USA

Liming Tao

4. Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

JiaBei Lin, Ronen Marmorstein & James Shorter

Contributions

X.Y. conceived and supervised the study. L.H. and X.Y. designed the experiments. L.H. performed most experiments. T.A. initiated the study and performed part of in vitro assays. G.Z. helped with in vitro assays. S.Y. helped with p53-related experiments. L.T. performed protein sequence analysis. R.M. helped with protein production in insect cells and advised on thermal shift assay. J.S. and J.L. provided HSP104 protein and advised on protein folding assays. X.Y. and L.H. prepared the manuscript with major inputs from J.S. and R.M. and comments from all other authors.

Corresponding author

Correspondence to [Xiaolu Yang](#).

Ethics declarations

Competing interests

X.Y. is a founder and equity holder of Evergreen Therapeutics LLC, which received investments from Wealth Strategy Holding Limited. J.S. is a consultant for Dewpoint Therapeutics, Maze Therapeutics, Vivid Sciences, Korro Bio, and ADRx. The other authors declare no competing financial interests.

Additional information

Peer review information *Nature* thanks Jerson Silva 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 Purification of recombinant DAXX proteins and their ability to prevent luciferase misfolding and aggregation.](#)

a, Scheme for purifying DAXX–6xHis from bacteria BL21(DE3) and insect Sf9 cells. To generate full-length protein, DAXX was fused to GST at the N terminus with a TEV protease cleavage site inserted in between GST and DAXX, and to 6xHis at the C terminus. The fusion protein was first purified with glutathione resins. Beads-bound GST–DAXX–6xHis protein was treated with TEV protease to released DAXX–6xHis, which was

subsequently purified with Ni-NTA resins. After elution with imidazole, DAXX-6xHis was further purified with ion-exchange and gel-filtration columns, and concentrated as needed. **b**, **c**, DAXX-6xHis purified from bacteria BL21(DE3) cells (**b**) and insect Sf9 cells (**c**) were analysed by SDS-PAGE and Coomassie blue staining. Bovine serum albumin (BSA) was used as a protein standard (**c**). Mass spectrometry analysis indicated that the vast majority of species in the minor bands of the DAXX preps were derived from DAXX. **d**, Scheme for purifying Flag-DAXX from HEK293T cells. Flag-DAXX was transiently transfected in HEK293T cells and purified using anti-Flag M2 beads. After elution with 3xFlag peptides, Flag-DAXX was further purified with ion-exchange and gel filtration columns, and concentrated as needed. **e**, Flag-DAXX purified from HEK293T cells was analysed by SDS-PAGE and Coomassie blue staining. Mass spectrometry analysis indicated that the vast majority of species in the minor bands of the DAXX preps were derived from DAXX. **f–h**, DAXX proteins purified from bacteria, insect cells, and mammalian cells protect luciferase from heat-induced inactivation and aggregation. Luciferase (**f**, **g**, 5 nM; **h**, 200 nM) was heated at 42 °C in the presence of indicated concentrations of GST, DAXX-6xHis (from bacteria) or HSP70 (plus HSP40 at a half concentration, same below) for 1 min (**f**), or in the presence or absence of GST, DAXX-6xHis (from bacteria), DAXX-6xHis (from insect cells), or Flag-DAXX (200 nM each) for the indicated times (**g**, **h**). Shown are luciferase activity relative to the native protein (**f**, **g**) and relatively turbidity measured at OD₆₀₀ (**h**). The DAXX protein purified from HEK293T cells appeared to be more active than those purified from bacteria and insect cells **i**, **j**. Protective activity of DAXX for a higher amount of luciferase. Luciferase (50 nM) was heated at 42 °C in the presence of the indicated concentrations of GST, DAXX-6xHis (from Sf9 cells), or HSP70 for 1 min (**i**), or in the presence of absence of GST, DAXX-6xHis (Sf9 cells) or HSP70-HSP40 (200 nM each) for the indicated times. Luciferase activity was normalized to native protein. RT-CTRL, control luciferase sample kept at room temperature. Assays in **b**, **c** and **e** have been performed three times with similar results. Numerical data are mean ± s.d. ($n = 3$) and are representative of three (**f**, **j**) or two (**g**, **h**, **i**) independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns, not significant; unpaired Student's *t*-test.

Source data

Extended Data Fig. 2 DAXX prevents α -Syn and A β 42 aggregation, acts independently of ATP, and probably exists as a monomer.

a, b, PFF-induced aggregation of soluble α -Syn monomers and its inhibition by DAXX. α -Syn monomers (13.3 μ M) was incubated alone, together with in α -Syn PFFs (133 nM) (**a**), or together with in α -Syn PFFs (133 nM) in the presence of GST (0.2 μ M), HSPs (0.2 μ M HSP70, 0.1 μ M HSP40, and 0.4 μ M HSP104(A503S)), and DAXX–6xHis (from Sf9 cells, 0.1, 0.2, and 0.4 μ M) (**b**). Aggregation was monitored by real-time quaking-induced conversion (RT-QuIC) assay. **c–e**, DAXX suppresses A β 42 fibrillization for a prolonged incubation, during which DAXX itself did not form fibrils or other sedimentable aggregates. A β 42 monomers (10 μ M) and DAXX–6xHis (from Sf9 cells, 0.05, 0.1, 0.2, 0.4 and 0.6 μ M) were incubated alone (**c, e**) or together (**d**) at 37 °C for 120 h. Formation of fibrils was analysed by ThT fluorescence assay (**c, d**). Solubility of DAXX was analysed by sedimentation assay (**e**). **f, g**, DAXX blocks A β 42 monomers to form PFFs that accelerate aggregation of fresh A β 42 monomers and A β 42 PFF-induced aggregation of fresh A β 42 monomers. A β 42 monomers (10 μ M) were incubated at 37 °C alone, together A β 42 PFFs (6 nM) (**f, g**), A β 42 (6 nM) preincubated with DAXX–6xHis (from Sf9 cells) at a 100:1 molar ratio (A β 42/DAXX), A β 42/DAXX plus DAXX–6xHis (0.6 μ M) (DAXX) (**f**), or A β 42 PFFs (6 nM) in the presence of DAXX (**g**). Formation of fibrils was analysed by ThT fluorescence assay. Assays in **f** and **g** were done at the same time. **h, i**, The chaperone activity of DAXX is not affected by the addition of ATP or the treatment of apyrase. Luciferase (0.2 μ M) was heated at 42 °C in presence of GST (0.2 μ M) (**h, i**), DAXX–6xHis (insect cells, 0.2 μ M) with or without ATP (5 mM ATP-Mg²⁺ plus an ATP-regeneration system) and apyrase as indicated (**h**), or HSP70–HSP40 (0.2 and 0.1 μ M, respectively) with or without apyrase (**i**). Aggregation formation was monitored by OD at 600 nm. **j**, DAXX does not bind to ATP. Recombinant DAXX–6xHis and HSP70 were incubated with agarose beads conjugated without ATP (–) or with ATP via the phosphate moiety (AP-ATP), ribose moiety (EDA-ATP), or the adenine base at different positions (6AH-ATP and 8AH-ATP). The input and pulldown samples were analysed

by western blot. **k**, DAXX exists as a homogeneous species of relatively low molecular weights. Recombinant Flag–DAXX protein was analysed by Superdex 200 10/300 GL column. Proteins standards (in kDa) are indicated. **l**, DAXX likely exists as a monomer. Recombinant Flag–DAXX (1 μ M) was crosslinked with indicated concentration of DSS at 25 °C for 30 min and analysed by western blot. Flag–p53 (1 μ M), which is expected to be a tetramer, was used as control. Similar results were obtained for DAXX–6xHis. Assays have been performed three (**b–e, k, l**) or two (**a, h, i, j**) times with similar results. Numerical data are mean \pm s.d. ($n = 3$) and are representative of three independent experiments (**f, g**).

Source data

Extended Data Fig. 3 DAXX can dissolve small luciferase aggregates, but not large luciferase aggregates or α -Syn fibrils.

a–c, DAXX dissolves and reactivates heat-denatured luciferase aggregates. Heat-denatured luciferase (5 nM) was incubated at 25 °C with GST (100 nM), DAXX–6xHis (from *E. coli*, 100 nM), and HSPs (100 nM HSP70, 50 nM HSP40 and 200 nM HSP104(A503S)) for the indicated times (**a**), or with the indicated amounts of GST or DAXX–6xHis for 90 min (**b, c**). Shown are luciferase activity relative to that of native luciferase (**b**), and the amount of luciferase in the SN after sedimentation (**a, c**). **d**, Disaggregase activity of DAXX proteins purified from different sources. Relative activity of heat-denatured luciferase (5 nM) that was incubated at 25 °C for 90 min with increasing concentrations of DAXX purified from the indicated sources. **e, f**, Disaggregase activity of DAXX for a higher amount of denatured luciferase. Relative activity of heat-denatured luciferase (50 nM) that was incubated at 25 °C with the increasing molar ratios of DAXX–6xHis (from Sf9 cells) for the indicated times (**e**) or 90 min (**f**). **g**, DAXX achieves the maximal recovery of luciferase activity at fivefold excess. Heat-denatured luciferase (0.1, 0.2, 0.5, 1 and 2 μ M) was incubated with DAXX–6xHis (0.1 μ M) at 25 °C for 90 min. Luciferase activity is relative to that of native luciferase. **h**, DAXX restores the native conformation to denatured luciferase. Native or heat-denatured luciferase (1 μ M) incubated alone or together with GST (1 μ M) or DAXX–6xHis (0.1 \times : 0.1 μ M; 0.5 \times : 0.5 μ M) for 90 min were examined by circular dichroism spectroscopy.

Data were analysed by CAPITO, with the percentages of β -strand shown in Fig. 2b. **i**, Different sizes of luciferase aggregates generated by heat and urea treatments. Luciferase (1 μ M) denatured by heat or urea was fractionated on Superdex 200 10/300 GL column. Fractions were analysed by western blot and the relative abundance of luciferase is indicated. **j**, DAXX cannot reactive urea-denatured luciferase. Relative activity of urea-denatured luciferase (5 nM) that was incubated with GST (0.2 μ M), DAXX (0.2 and 1 μ M), or HSPs (0.2 μ M HSP70, 0.1 μ M HSP40 and 0.4 μ M HSP104(A503S)) at 25 °C for 90 min. **k, l**, Luciferase denatured by heat (**k**) or urea (**l**) was fractionated on gel filtration chromatography. Fractions in the range of 44 to 2,000 kDa were incubated with buffer, lysozyme (0.1 μ M), DAXX–6xHis (0.1 μ M), or HSP70–HSP40–HSP104(A503S) (0.1, 0.05 and 0.2 μ M, respectively) at 25 °C for 90 min, and luciferase activity was determined. **m–p**, DAXX is unable to dissolve α -Syn fibrils. Preformed α -Syn fibrils (0.2 μ M) were treated with GST (0.2 μ M) (**m–p**), DAXX–6xHis at the indicated molar ratios (**m, n**) or at 0.2 μ M (**o, p**), HSPs (0.2 μ M HSP70, 0.1 μ M HSP40, and 0.4 μ M HSP104(A503S) plus ATP and an ATP regeneration system) (**m–p**), or both DAXX and HSPs (**o, p**). Reactions mixtures were analysed by dot blot (**m, o**), and soluble α -Syn relative to total α -Syn was quantified (**n, p**) ($n = 3$). Assays in **a, c, h, i, k–p** have been performed three times with similar results. Numerical data are mean \pm s.d. ($n = 3$) and are representative of three independent experiments (**b, d–g, j**). ** $P < 0.01$, *** $P < 0.001$; unpaired Student's *t*-test.

Source data

Extended Data Fig. 4 DAXX unfolds misfolded LucD and protects against protein aggregation and oligomerization in cells.

a, A schematic representation of compact misfolded LucD monomers, the unfolded intermediates, and the native conformers, as well as their sensitivity to brief trypsin digestion. **b**, DAXX changes trypsin sensitivity of LucD. LucD (50 nM) was incubated with DAXX–xHis (100 nM), GST (100 nM), or HSPs (100 nM HSP70, 50 nM HSP40 and 200 nM HSP104(A503S)) at 25 °C. At indicated time points, aliquots of luciferase were incubated with 2.5 μ M trypsin at 22 °C for 2 min and were analysed

by western blot. Shown is luciferase band intensity. A representative western blot is presented in Fig. 2g. **c**, DAXX increases the enzymatic activity of LucD. Misfolded LucD monomers (50 nM) were incubated with GST (100 nM) or DAXX–6xHis (100 nM) for indicated durations and assayed for luciferase activity. **d**, DAXX elevates the levels of nLucDM, but not nLuc, in cells. nLuc or nLucDM was transfected together with empty vector (EV) or Flag–DAXX in HEK293T cells. Cell lysates were analysed by western blot 24 h after transfection. **e, f**, DAXX reduces aggregation, but not expression, of ATXN1(82Q) in cells. U2OS cells transfected with HA-Atxn1-82Q together with EV, Flag–DAXX (**e, f**) or GFP–HSP70 (**f**) were analysed by western blot (**e**), or immunofluorescence with anti-HA (red) and anti-Flag (green) antibodies (**f**; scale bar, 10 μ m). Quantification of the percentage of cells containing different sizes Atxn1-82Q inclusions is shown in Fig. 2i. **g**, Schematic representation of BiFC assay based on Venus, an improved version of yellow fluorescent protein (YFP). **h–j**, DAXX inhibits α -Syn oligomerization in cells. HEK293T cells were transfected with V1S and SV2 individually, or together with empty vector (EV) or DAXX. Cells were analysed by western blot for protein expression (**h**) and by fluorescence microscopy for BiFC signals and Flag–DAXX expression (**i**; scale bars, 100 μ m), with the quantification of BiFC signals shown in **j**. Assay in **d, e, f, h** and **i** have been performed two times with similar results. Numerical data are mean \pm s.d. ($n = 3$) and are representative of three independent experiments (**b, c, j**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; unpaired Student's *t*-test.

Source data

Extended Data Fig. 5 DAXX binds to misfolded proteins and depends on the polyD/E region for its activity.

a, The DAXX-luciferase interaction is increased upon heat shock. HEK293T cells were transfected with HA–DAXX and Flag–nLuc as indicated, and treated with or without heat shock. Cell lysates were immunoprecipitated with anti-Flag mAb (M2) beads. Input and immunoprecipitated samples were analysed by western blot. **b**, DAXX preferentially binds to heat-denatured luciferase. GST or GST–DAXX (100 nM each) was incubated with native (N) or heat-denatured (D) 6xHis-

luciferase immobilized on Ni-NTA agarose. The input and pulldown samples were analysed by western blot. **c**, Binding of DAXX to cellulose-bound peptide scans derived from luciferase, p53, MDM2, H3.3, H4 and DAXX. Each peptide contained thirteen amino acid residues that overlapped adjacent peptides by ten. **d**, Protein sequence of DAXX. The poly D/E region is marked in red colour, and the four continuous runs of Asp/Glu (with 14, 11, 7, and 5 residues, respectively) are underlined. **e**, Schematic presentation of full-length DAXX and its mutants. DAXX(D/E) was fused to GST at the N terminus for protein stabilization. **f**, DAXX(Δ D/E) and DAXX(D/E) lack unfoldase activity. Misfolded LucD monomers (50 nM) were incubated with DAXX(Δ D/E) or DAXX(D/E) (100 nM each) for the indicated times and assayed for luciferase activity (mean \pm s.d., $n = 3$). **g**, **h**, DAXX(Δ D/E) and DAXX(D/E) remain soluble during heat treatment. Recombinant DAXX Δ D/E (**g**) and D/E (**h**) proteins were heated at 42 °C for the indicated durations. Luciferase (200 nM) was used as a positive control. Aggregation formation was monitored by OD₆₀₀ and normalized to the luciferase control. Assays in **a–c** have been performed two times with similar results. Numerical data are mean or mean \pm s.d. ($n = 3$) and are representative of two independent experiments (**f–h**).

[Source data](#)

Extended Data Fig. 6 Role of other polyD/E proteins in protein folding.

a, Sequences of human SET and ANP32A proteins. The poly D/E region is marked in red colour, and the continuous runs of Asp/Glu are underlined. **b**, **c**, ANP32A and SET does not block α -Syn fibrillization. α -Syn monomers (70 μ M) were incubated with SET or ANP32A (0.4 μ M each) for 7 days. Samples were analysed by electron microscopy (**b**) and ThT staining (**c**). Scale bar, 100 nm. **d**, SET and ANP32A are unable to solubilize urea-denatured luciferase. Urea-denatured luciferase (5 nM) was incubated with GST, SET, ANP32A (0.2 μ M each) or HSPs (0.2 μ M HSP70, 0.1 μ M HSP40, and 0.4 μ M HSP104(A503S)) at 25 °C for 90 min. Activity relative to that of the native control are shown. **e**, **f**, Unfoldase activity of ANP32A and SET. Misfolded LucD (50 nM) was incubated with GST, SET or

ANP32A (200 nM each) at 25 °C. At indicated time points, aliquots of refolding luciferase were incubated with 2.5 μM trypsin at 22 °C for 2 min, denatured in SDS sample buffer, and analysed by western blot (**e**), with the quantification showed in (**f**). **g, h**, Schematic presentation of SET and its deletion mutants (**g**), and the numbers of Asp (D) and Glu (E) in each mutant (**h**). **i**, Heat-inactivated luciferase (5 nM) was incubated at 25 °C with SET or its deletion mutants (200 nM each) for 90 min. Activity relative to that of native luciferase are shown. **j**, Number of polyD/E proteins different species. **k, l**, Gene Ontology analysis of polyD/E proteins in humans. Proteins are classified into pie chart based on their molecular functions (**k**) and protein classes (**l**). Assays have been performed two (**b**) or three (**e**) times with similar results. Numerical data are mean ± s.d. ($n = 3$) and are representative of two independent experiments (**c, d, f, i**).

** $P < 0.01$, unpaired Student's *t*-test.

[Source data](#)

[**Extended Data Fig. 7 DAXX maintains the native conformation of both p53 and MDM2.**](#)

a, DAXX abrogates p53 fibrillization. Recombinant wild-type p53 and DAXX–6xHis proteins (5 μM each) were incubated alone or together at 37 °C for 2 h in the presence of ThT (25 μM). Formation of amyloid fibrils was assayed by ThT. **b, c**, DAXX(ΔD/E) and DAXX(D/E) cannot protect p53 from aggregation. Native p53 (n-p53) (**b**), or denatured p53 aggregates (d-p53) (**c**), (100 nM each) was incubated with GST, Flag–DAXX, DAXX(D/E) or DAXX(ΔD/E) (200 nM each) at 37 °C (**b**) or at 25 °C (**c**) for the indicated times. Samples were partitioned into supernatant (soluble) and pellet (insoluble) fractions via sedimentation, and analysed by western blot. As in Fig. 4a, b except that DAXX(D/E) and DAXX(ΔD/E) samples are included. **d, e, g, h**, DAXX restore native conformation to denatured p53 and MDM2. Native p53 (**d**) or native MDM2 (n-MDM2, **g**), or denatured p53 (**e**) or MDM2 (d-MDM2, **h**), (1 μM each) was incubated alone or together with GST or DAXX–6xHis (0.5, 1, 2 μM, from Sf9 cells) at the indicated molar ratios for 3 h and analysed by thermal shift assay. The transition of the unfolding curve represents the temperature at which the protein unfolding occurs (T_m). **f**, DAXX dissolve preformed MDM2

aggregates. d-MDM2 (100 nM) was incubated with Flag–DAXX (200 nM) at 25 °C for the indicated times. Supernatant (soluble) and pellet (insoluble) fractions after sedimentation were analysed by western blot. **i**, DAXX enhances MDM2-mediated p53 ubiquitination. Native p53 (20 nM) was incubated with native MDM2 (45 nM) in the presence or absence of DAXX (20 or 100 nM) at 37 °C for 1.5 h. E1, E2 and His-ubiquitin (His-Ub) were then added for in vitro ubiquitination assay. The reaction mixtures were analysed by western blot. **j**, Native MDM2-mediated ubiquitination of native p53 (20 nM) in the presence or absence of Flag–DAXX (100 nM), or of denatured p53 (20 nM) pre-incubated with or without Flag–DAXX (100 nM) for 3 h at 25 °C. **k**, **l**, DAXX reduces p53 levels in cells, but does not alter the largely diffuse nuclear localization pattern of p53. Flag–p53 was transfected into U2OS cells together with empty vector or DAXX. Cells were analysed by immunofluorescence (**k**) and western blot (**l**). **m–o**, H1299 cells inducibly expressing wild-type p53 or p53(R175H) were transfected with control vector (–) or HA–DAXX. Upon induction of p53 expression by Dox (1 µg ml^{–1}), cells were analysed for protein levels by western blot with relative p53/GAPDH ratios indicated (**m**) and for mRNA levels of *p53* (**n**) and p53 target genes (**o**) by qRT–PCR. Scale bar, 10 µm. Assays in panels have been performed two (**d**, **e**, **g**, **h**, **k–m**) or three (**b**, **c**, **f**, **i**, **j**) times with similar results. Numerical data are mean ± s.d. (*n* = 3) and are representative of two independent experiments (**a**, **n**, **o**). **P* < 0.05, ***P* < 0.01, ns, not significant; unpaired Student’s *t*-test.

[Source data](#)

[Extended Data Fig. 8 DAXX restores native conformation and function of mutant p53.](#)

a, DAXX prevents p53(R175H) aggregation. p53(R175H) protein (100 nM) was incubated with GST or Flag–DAXX (200 nM each) at 37 °C for the indicated times. SN and pellet fractions were analysed by western blot. **b**, DAXX blocks p53(R175H) PFF-induced fibrillization of p53. Wild-type p53 (5 µM) was incubated with or without p53(R175H) PFFs and DAXX as indicated. Fibril formation was assayed by ThT binding. **c**, DAXX reduces p53(R175H) aggregates in cells. Flag–p53(R175H) was transfected into the U2OS cells together with empty vector (Ctrl) or HA–DAXX. Cells were

analysed by immunofluorescence. Scale bar, 10 µm. Part of the images are also shown in Fig. 4*i*. **d**, DAXX partially restores the function of mutant p53. H1299 cells inducibly expressing p53(R175H) were transfected with control vector (–) or HA-DAXX. Upon induction of p53 expression by Dox (1 µg ml^{–1}), cells were analysed for the expression of p53 target genes by RT–PCR. **e–g**, Effect of DAXX on aggregation of endogenous mutant p53. MDA-MB-231 cells were transduced with lentiviral vectors expressing control or DAXX shRNA (**e, f**), or transfected with control siRNA, *DAXX* siRNA, and/or an siRNA-resistant form of DAXX (Flag-DAXX) as indicated (**g**). Cells were immunostained with anti-p53 (DO-1) and anti-fibrillar oligomer (A11) antibodies (**e, g**) and quantified (**f**). Scale bar, 50 µm. **h–j**, Knocking down DAXX enhances growth and tumorigenicity of MDA-MB-231 cells. Control and DAXX-knockdown MDA-MB-231 cells were assayed for adherent proliferation, protein expression (**h**), and soft-agar colony formation (21 days), with number and sizes of colonies (**i**) and representative images of colonies (**j**) shown. **k, l**, Overexpressing DAXX inhibits growth and tumorigenicity of MDA-MB-231 cells. MDA-MB-231 transduced with pBabe or pBabe-Flag-DAXX were assay for adherent proliferation for 5 days (**k**) and soft-agar colony formation (21 days), with representative images of colonies shown (**l**). Assays have been performed two (**a**) or three (**c, e, g**) times with similar results. Numerical data are mean ± s.d. ($n = 3$ for **b, d**, and 6 for **i, h, k**) and are representative of three independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; unpaired Student's *t*-test.

[Source data](#)

Extended Data Table 1 Human polyD/E proteins containing 35 or more D or E residues in any 50-residue window

Supplementary information

[Supplementary Information](#)

This file contains legends for Supplementary Figure 1 and Supplementary Table 1.

Reporting Summary

Supplementary Figure 1

Unprocessed western blots and/or gels associated with the data presented in the Figures and Extended Data Figures.

Supplementary Table 1

Human polyD/E proteins containing 35 or more D or E residues in any 50-residue window.

Source data

[**Source Data Fig. 1**](#)

[**Source Data Fig. 2**](#)

[**Source Data Fig. 3**](#)

[**Source Data Fig. 4**](#)

[**Source Data Extended Data Fig. 1**](#)

[**Source Data Extended Data Fig. 2**](#)

[**Source Data Extended Data Fig. 3**](#)

[**Source Data Extended Data Fig. 4**](#)

[**Source Data Extended Data Fig. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Huang, L., Agrawal, T., Zhu, G. *et al.* DAXX represents a new type of protein-folding enabler. *Nature* **597**, 132–137 (2021).
<https://doi.org/10.1038/s41586-021-03824-5>

- Received: 21 December 2019
- Accepted: 15 July 2021
- Published: 18 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03824-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03824-5>

Amendments & Corrections

- [**Author Correction: Commensal-driven immune zonation of the liver promotes host defence**](#) [09 August 2021]
Author Correction •
- [**Publisher Correction: Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice**](#) [03 August 2021]
Publisher Correction •
- [**Addendum: The mutational constraint spectrum quantified from variation in 141,456 humans**](#) [09 August 2021]
Addendum •

- Author Correction
- [Published: 09 August 2021](#)

Author Correction: Commensal-driven immune zonation of the liver promotes host defence

- [Anita Gola](#) [ORCID: orcid.org/0000-0003-1431-1398](#)^{1 nAff9},
- [Michael G. Dorrington](#)^{2 na1},
- [Emily Speranza](#)^{1,3 na1},
- [Claudia Sala](#) [ORCID: orcid.org/0000-0002-4889-1047](#)⁴,
- [Rochelle M. Shih](#) [ORCID: orcid.org/0000-0002-3218-4479](#)¹,
- [Andrea J. Radtke](#)¹,
- [Harikesh S. Wong](#)¹,
- [Antonio P. Baptista](#)^{1,5,6},
- [Jonathan M. Hernandez](#)⁷,
- [Gastone Castellani](#)⁸,
- [Iain D. C. Fraser](#)² &
- [Ronald N. Germain](#) [ORCID: orcid.org/0000-0003-1495-9143](#)¹

[Nature](#) volume **597**, page E1 (2021)

- 500 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Chemokines](#)

- [Imaging the immune system](#)

The [Original Article](#) was published on 25 November 2020

A Correction to this paper has been published:
<https://doi.org/10.1038/s41586-021-03346-0>.

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-020-2977-2> Published online 25 November 2020

In Fig. 1b, Extended Data Fig. 3e and Supplementary Table 1 of this Article, an enzyme was inadvertently mislabelled as ALS (acetolactate synthase) instead of ARG1 (arginase 1). The figures and corresponding captions have been corrected and the Supplementary Information has been replaced online. The Author(s), under exclusive licence to Springer Nature Limited 2021

Author information

Author notes

1. Anita Gola

Present address: Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, NY, USA

2. These authors contributed equally: Michael G. Dorrington, Emily Speranza

Affiliations

1. Lymphocyte Biology Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Anita Gola, Emily Speranza, Rochelle M. Shih, Andrea J. Radtke, Harikesh S. Wong, Antonio P. Baptista & Ronald N. Germain

2. Signaling Systems Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Michael G. Dorrington & Iain D. C. Fraser

3. Innate Immunity and Pathogenesis Section, Laboratory of Virology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Emily Speranza

4. Department of Physics and Astronomy, University of Bologna, Bologna, Italy

Claudia Sala

5. Laboratory of Immunoregulation and Mucosal Immunology, VIB-UGent Center for Inflammation Research, Ghent, Belgium

Antonio P. Baptista

6. Department of Internal Medicine and Pediatrics, Ghent University, Ghent, Belgium

Antonio P. Baptista

7. Metastasis Biology Section, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Jonathan M. Hernandez

8. Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy

Gastone Castellani

Corresponding authors

Correspondence to [Anita Gola](#) or [Ronald N. Germain](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Gola, A., Dorrington, M.G., Speranza, E. *et al.* Author Correction: Commensal-driven immune zonation of the liver promotes host defence. *Nature* **597**, E1 (2021). <https://doi.org/10.1038/s41586-021-03346-0>

- Published: 09 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03346-0>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03346-0>

- Publisher Correction
- [Published: 03 August 2021](#)

Publisher Correction: Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice

- [Raoul De Gasparo^{1 na1}](#),
- [Mattia Pedotti^{1 na1}](#),
- [Luca Simonelli¹](#),
- [Petr Nickl](#) [ORCID: orcid.org/0000-0002-4289-0502²](#),
- [Frauke Muecksch](#) [ORCID: orcid.org/0000-0002-0132-5101³](#),
- [Irene Cassaniti⁴](#),
- [Elena Percivalle⁴](#),
- [Julio C. C. Lorenzi⁵](#),
- [Federica Mazzola¹](#),
- [Davide Magrì](#) [ORCID: orcid.org/0000-0003-3790-5332⁶](#),
- [Tereza Michalcikova²](#),
- [Jan Haviernik⁷](#),
- [Vaclav Honig](#) [ORCID: orcid.org/0000-0003-0469-4604^{7,8}](#),
- [Blanka Mrazkova²](#),
- [Natalie Polakova²](#),
- [Andrea Fortova⁷](#),
- [Jolana Tureckova²](#),
- [Veronika Iatsiuk²](#),
- [Salvatore Di Girolamo¹](#),
- [Martin Palus^{7,8}](#),
- [Dagmar Zudova²](#),

- [Petr Bednar^{7,9}](#),
- [Ivana Bukova²](#),
- [Filippo Bianchini¹](#),
- [Dora Mehn⁶](#),
- [Radim Nencka¹⁰](#),
- [Petra Strakova⁷](#),
- [Oto Pavlis¹¹](#),
- [Jan Rozman](#) [ORCID: orcid.org/0000-0002-8035-8904²](#),
- [Sabrina Gioria⁶](#),
- [Josè Camilla Sammartino⁴](#),
- [Federica Giardina⁴](#),
- [Stefano Gaiarsa⁴](#),
- [Qiang Pan-Hammarström](#) [ORCID: orcid.org/0000-0003-1990-8804¹²](#),
- [Christopher O. Barnes¹³](#),
- [Pamela J. Bjorkman](#) [ORCID: orcid.org/0000-0002-2277-3990¹³](#),
- [Luigi Calzolai](#) [ORCID: orcid.org/0000-0002-8474-7974⁶](#),
- [Antonio Piralla](#) [ORCID: orcid.org/0000-0002-6062-2579⁴](#),
- [Fausto Baldanti^{4,14}](#),
- [Michel C. Nussenzweig^{5,15}](#),
- [Paul D. Bieniasz](#) [ORCID: orcid.org/0000-0002-2368-3719^{3,15}](#),
- [Theodora Hatzioannou³](#),
- [Jan Prochazka²](#),
- [Radislav Sedlacek](#) [ORCID: orcid.org/0000-0002-3352-392X²](#),
- [Davide F. Robbiani](#) [ORCID: orcid.org/0000-0001-7379-3484¹](#),
- [Daniel Ruzek](#) [ORCID: orcid.org/0000-0003-4655-2380^{7,8}](#) &
- [Luca Varani](#) [ORCID: orcid.org/0000-0002-0963-0987¹](#)

Nature volume 597, page E2 (2021)

- 500 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Antibody therapy](#)
- [Immunization](#)
- [Protein design](#)
- [SARS-CoV-2](#)
- [Viral infection](#)

The [Original Article](#) was published on 25 March 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03461-y> published online 25 March 2021

In this Article, authors Raoul De Gasparo and Mattia Pedotti should not be linked to affiliation 15, ‘Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA’ and should instead be linked to the equal contribution footnote, 16. In addition, affiliation 14 should read ‘Università degli Studi di Pavia, Pavia, Italy’. The Article has been corrected online.

Author information

Author notes

1. These authors contributed equally: Raoul De Gasparo, Mattia Pedotti

Affiliations

1. Institute for Research in Biomedicine, Università della Svizzera italiana (USI), Bellinzona, Switzerland

Raoul De Gasparo, Mattia Pedotti, Luca Simonelli, Federica Mazzola, Salvatore Di Girolamo, Filippo Bianchini, Davide F. Robbiani & Luca Varani

2. Czech Centre of Phenogenomics and Laboratory of Transgenic Models of Diseases, Institute of Molecular Genetics of the Czech Academy of Sciences, Vestec, Czech Republic

Petr Nickl, Tereza Michalcikova, Blanka Mrazkova, Natalie Polakova, Jolana Tureckova, Veronika Iatsiuk, Dagmar Zudova, Ivana Bukova, Jan Rozman, Jan Prochazka & Radislav Sedlacek

3. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

Frauke Muecksch, Paul D. Bieniasz & Theodora Hatzioannou

4. Molecular Virology Unit, Microbiology and Virology Department, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

Irene Cassaniti, Elena Percivalle, Josè Camilla Sammartino, Federica Giardina, Stefano Gaiarsa, Antonio Piralla & Fausto Baldanti

5. Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA

Julio C. C. Lorenzi & Michel C. Nussenzweig

6. European Commission, Joint Research Centre, Ispra, Italy

Davide Magri, Dora Mehn, Sabrina Gioria & Luigi Calzolai

7. Veterinary Research Institute, Brno, Czech Republic

Jan Haviernik, Vaclav Honig, Andrea Fortova, Martin Palus, Petr Bednar, Petra Strakova & Daniel Ruzek

8. Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic

Vaclav Honig, Martin Palus & Daniel Ruzek

9. Faculty of Science, University of South Bohemia, Ceske Budejovice,
Czech Republic

Petr Bednar

10. Institute of Organic Chemistry and Biochemistry of the Czech
Academy of Sciences, Prague, Czech Republic

Radim Nencka

11. Center of Biological Defense, Military Health Institute, Military
Medical Agency, Techonin, Czech Republic

Oto Pavlis

12. Department of Biosciences and Nutrition, Karolinska Institutet,
Huddinge, Sweden

Qiang Pan-Hammarström

13. Division of Biology and Biological Engineering, California Institute of
Technology, Pasadena, CA, USA

Christopher O. Barnes & Pamela J. Bjorkman

14. Department of Clinical Surgical Diagnostic and Pediatric Sciences,
Università degli Studi di Pavia, Pavia, Italy

Fausto Baldanti

15. Howard Hughes Medical Institute, The Rockefeller University, New
York, NY, USA

Michel C. Nussenzweig & Paul D. Bieniasz

Corresponding authors

Correspondence to [Davide F. Robbiani](#) or [Daniel Ruzek](#) or [Luca Varani](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

De Gasparo, R., Pedotti, M., Simonelli, L. *et al.* Publisher Correction: Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice. *Nature* **597**, E2 (2021). <https://doi.org/10.1038/s41586-021-03719-5>

- Published: 03 August 2021
- Issue Date: 02 September 2021
- DOI: <https://doi.org/10.1038/s41586-021-03719-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03719-5>

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

- Addendum
- Open Access
- [Published: 09 August 2021](#)

Addendum: The mutational constraint spectrum quantified from variation in 141,456 humans

- [Sanna Gudmundsson](#)^{1,2,15},
- [Konrad J. Karczewski](#) ORCID: orcid.org/0000-0003-2878-4671^{1,2},
- [Laurent C. Francioli](#) ORCID: orcid.org/0000-0002-3039-8895^{1,2},
- [Grace Tiao](#)^{1,2},
- [Beryl B. Cummings](#)^{1,2,3},
- [Jessica Alföldi](#) ORCID: orcid.org/0000-0001-9713-6200^{1,2},
- [Qingbo Wang](#) ORCID: orcid.org/0000-0002-9110-5830^{1,2,4},
- [Ryan L. Collins](#) ORCID: orcid.org/0000-0003-1268-9995^{1,4,5},
- [Kristen M. Laricchia](#)^{1,2},
- [Andrea Ganna](#)^{1,2,6},
- [Daniel P. Birnbaum](#)^{1,2},
- [Laura D. Gauthier](#)⁷,
- [Harrison Brand](#)^{1,5},
- [Matthew Solomonson](#)^{1,2},
- [Nicholas A. Watts](#)^{1,2},
- [Daniel Rhodes](#)⁸,
- [Moriel Singer-Berk](#)^{1,2},
- [Eleina M. England](#)^{1,2},
- [Eleanor G. Seaby](#)^{1,2},
- [Jack A. Kosmicki](#)^{1,2,4},
- [Raymond K. Walters](#) ORCID: orcid.org/0000-0001-8422-6530^{1,2,9},
- [Katherine Tashman](#)^{1,2,9},
- [Yossi Farjoun](#) ORCID: orcid.org/0000-0002-7002-2868⁷,
- [Eric Banks](#)⁷,
- [Timothy Poterba](#)^{1,2,9},
- [Arcturus Wang](#)^{1,2,9},

- [Cotton Seed](#)^{1,2,9},
- [Nicola Whiffin](#)^{1,2,10,11},
- [Jessica X. Chong](#)¹²,
- [Kaitlin E. Samocha](#) ORCID: orcid.org/0000-0002-1704-3352¹³,
- [Emma Pierce-Hoffman](#)^{1,2},
- [Zachary Zappala](#)^{1,2,14},
- [Anne H. O'Donnell-Luria](#) ORCID: orcid.org/0000-0001-6418-9592^{1,2,15,16},
- [Eric Vallabh Minikel](#)¹,
- [Ben Weisburd](#)⁷,
- [Monkol Lek](#)¹⁷,
- [James S. Ware](#) ORCID: orcid.org/0000-0002-6110-5880^{1,10,11},
- [Christopher Vittal](#)^{2,9},
- [Irina M. Armean](#) ORCID: orcid.org/0000-0002-1043-4756^{1,2},
- [Louis Bergelson](#)⁷,
- [Kristian Cibulskis](#)⁷,
- [Kristen M. Connolly](#)¹⁸,
- [Miguel Covarrubias](#)⁷,
- [Stacey Donnelly](#)¹,
- [Steven Ferriera](#)¹⁸,
- [Stacey Gabriel](#)¹⁸,
- [Jeff Gentry](#)⁷,
- [Namrata Gupta](#)^{1,18},
- [Thibault Jeandet](#)⁷,
- [Diane Kaplan](#)⁷,
- [Christopher Llanwarne](#)⁷,
- [Ruchi Munshi](#)⁷,
- [Sam Novod](#)⁷,
- [Nikelle Petrillo](#)⁷,
- [David Roazen](#)⁷,
- [Valentin Ruano-Rubio](#)⁷,
- [Andrea Saltzman](#)¹,
- [Molly Schleicher](#)¹,
- [Jose Soto](#)⁷,
- [Kathleen Tibbetts](#)⁷,
- [Charlotte Tolonen](#)⁷,
- [Gordon Wade](#)⁷,
- [Michael E. Talkowski](#) ORCID: orcid.org/0000-0003-2889-0992^{1,5,19},
- [Genome Aggregation Database Consortium](#),
- [Benjamin M. Neale](#) ORCID: orcid.org/0000-0003-1513-6077^{1,2,9},

- [Mark J. Daly^{1,2,6,9}](#) &
- [Daniel G. MacArthur^{1,2}](#) [ORCID: orcid.org/0000-0002-5771-2290^{1,2}](#) [nAff150](#) [nAff151](#)

[Nature](#) volume **597**, pages E3–E4 (2021)

- 918 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Medical genomics](#)
- [Rare variants](#)

The [Original Article](#) was published on 27 May 2020

[Download PDF](#)

Addendum to: *Nature* <https://doi.org/10.1038/s41586-020-2308-7> Published online 27 May 2020

This analysis explores the extent of loss-of-function (LoF) tolerance in human disease genes.

Databases of human population genetic variation, such as the Genome Aggregation Database (gnomAD), are generally expected to be depleted for variation with severe effects on health. As such, it is expected that genes that carry highly disruptive changes, predicted (p)LoF variants, in these databases are less likely to be responsible for severe human disease. However, the precise relationship between pLoF tolerance and human disease causation is not well-characterized.

In our Article, we reported a total of 2,636 variants in 1,815 genes that were homozygous in at least one individual and annotated as pLoF after applying both automated filtering and manual curation of both sequencing quality and functional annotation. We labelled these genes as ‘LoF-tolerant’, indicating that total functional loss of these genes appears to be compatible with life. This does not exclude the involvement of these genes in diseases compatible with presence in individuals in gnomAD¹. Neither the ‘LoF Transcript Effect Estimator’ (LOFTEE) nor manual curation took previous gene–phenotype associations into account, as this would create

a bias that affects downstream analyses and also may result in the spurious exclusion of true LoF-tolerant genes owing to previous false-positive reported associations with disease. This unbiased approach is appropriate for permitting downstream analyses, but it means that the enrichment of pLoF artefacts will remain higher in genes for which genetic disruption is genuinely associated with severe disease.

Prompted by comments on our original Article, we explored the degree to which our LoF-tolerant list includes genes associated with disease by manually curating the 158 genes (with 217 pLoF variants) on the LoF-tolerant list associated with autosomal recessive and X-linked traits in ‘Online Mendelian Inheritance in Man’ (OMIM) by an additional biocurator¹.

Of these genes, 71% ($n = 112$) are associated with phenotypes that are likely to be found in gnomAD, on the basis of gnomAD inclusion criteria. These are phenotypes such as infertility, hearing or visual impairment, benign or mild metabolic or haematological phenotypes, expected at similar frequency as the general population (95 phenotypes) and, to a lesser extent, traits that are likely to be depleted from gnomAD, but for which someone with the condition may participate in a common disease study (17 phenotypes). We observed an overrepresentation of traits that are likely to be found (60% versus 33%) and an underrepresentation of traits that are not expected to be found (29% versus 53%) in gnomAD (early-onset severe or lethal rare disease that generally would restrict participation in genetic studies) versus a control set of 100 random selected autosomal recessive and X-linked OMIM traits ($P = 3.0 \times 10^{-5}$, Fisher’s exact test) (Fig. 1a). We performed a thorough literature review of the 46 phenotypes that were initially not expected to be found in gnomAD, which revealed that 35% (16 out of 46) can be explained by evidence of mechanism of disease not being LoF ($n = 2$), variable expressivity ($n = 5$) or penetrance ($n = 3$), phenotype being responsive to treatment ($n = 4$) and onset after age of the individual in gnomAD ($n = 2$) (Fig. 1b, blue).

Fig. 1: Assessment of pLoF variants in LoF-tolerant genes associated with autosomal recessive and X-linked phenotypes in OMIM.

 **figure1**

a, Autosomal recessive and X-linked (AR) OMIM phenotypes: likely to be found (blue), likely to be depleted (yellow) or not expected (red) to be found in gnomAD, for the 158 phenotypes associated with LoF-tolerant genes in gnomAD and a set of 100 randomly selected AR and X-linked OMIM traits. *** $P = 3.0 \times 10^{-5}$, Fisher's exact test. **b**, Extended literature review of the 46 out of 158 OMIM phenotypes not expected to be found in gnomAD. **c**, Extended variant curation of 32 pLoF variants in 30 LoF-tolerant genes beyond criteria presented in our original Article revealed pLoF with suggested evasion of pLoF (purple), and pLoF with no conclusive (pink) or no evidence (grey) contradicting pLoF in these genes. NMD, nonsense-mediated decay. Further details are provided in [Supplementary Table 1](#).

In contrast to what is expected to be found in gnomAD, 32 pLoF variants are in 30 genes for which homozygous LoF has been associated with severe or lethal phenotypes in OMIM. However, 10 of these 30 genes had a limited number of cases reported ($n = 7$) or no reported biallelic LoF variants in humans ($n = 3$) (Fig. [1b](#), light red) and only 5 genes meet current ClinGen standards for a known LoF mechanism².

We evaluated the 32 variants by applying more stringent criteria, and identified several cases in which a variety of mechanisms may result in an evasion of true loss of gene function. For 15 variants, we found evidence that disputed our previous prediction (Fig. 1c, purple), including variants that are suspected to escape nonsense-mediated decay but that did not meet the criteria for rescue applied in our original Article ($n = 12$), one variant that was within a small homopolymer and thus is more likely to represent a sequencing error, one alignment error, and one variant that is in an overprinted transcript and is more probably a synonymous variant in the most biologically relevant transcript. For the 17 variants for which we cannot identify conclusive ($n = 9$) (Fig. 1c, pink) or any ($n = 8$) (Fig. 1c, grey) evidence for evasion of pLoF, there are several explanations that even our stringent curation cannot confidently exclude: for example, sample swaps, a variety of residual sequencing and annotation artefact classes, the presence of an individual in gnomAD who does actually have the expected phenotype, or simply variable expressivity, late age of onset or reduced penetrance of the disease phenotype itself. Further details regarding variant curation are available in [Supplementary Table 1](#) and from <https://gnomad.broadinstitute.org/downloads>, or the curation data can be viewed at the respective gene page at <https://gnomad.broadinstitute.org>.

In summary, this result emphasizes the well-established need for extremely careful curation of any pLoF variant observed in a population database such as gnomAD, especially for genes for which such variants are expected to be deleterious. The variants curated here are found at low frequency and are enriched for both sequencing and annotation errors^{3,4}. This enrichment is expected to be even larger in genes for which inactivation is associated with severe disease, because sequencing and annotation artefacts are distributed approximately uniformly across the genome, whereas true LoF variation is depleted in genes in which it results in a more detrimental effect. Although the pLoF variants found in the gnomAD dataset have been subjected to thorough quality control, any filtration process other than comprehensive experimental validation is insufficient to remove all artefacts.

In conclusion, population databases such as gnomAD are a powerful source of information when predicting human tolerance towards gene disruption. The list of LoF-tolerant genes identified in gnomAD is a useful class for downstream analysis that appears to largely comprise genes for which true homozygous disruption does not cause severe early-onset disease.

Authors S.G. and M.S.-B. carried out the analysis described in this Addendum. K.J.K., A.O.-L. and D.G.M. contributed to the experimental design, and A.O.-L. and D.G.M. supervised the work. S.G., M.S.-B., K.J.K., A.O.-L. and D.G.M. wrote the Addendum. A.O.-L. and D.G.M. contributed equally to this work.

We thank C. Arnoult, P. Ray and N. Thierry-Mieg for presenting the opportunity to further clarify the term LoF tolerance.

Supplementary Information is available in the online version of this Amendment.

References

1. 1.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).

2. 2.

Abou Tayoun, A. N. et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* **39**, 1517–1524 (2018).

3. 3.

MacArthur, D. G. & Tyler-Smith, C. Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* **19**, R125–R130 (2010).

4. 4.

MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

Author information

Author notes

1. Daniel G. MacArthur

Present address: Centre for Population Genomics, Garvan Institute of Medical Research, and UNSW Sydney, Sydney, New South Wales, Australia

2. Daniel G. MacArthur

Present address: Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia

3. Deceased: Pamela Sklar

Affiliations

1. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Sanna Gudmundsson, Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alfoldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, James S. Ware, Irina M. Armean, Stacey Donnelly, Namrata Gupta, Andrea Saltzman, Molly Schleicher, Michael E. Talkowski, Mark J. Daly, Patrick T. Ellinor, Tõnu Esko, Jose Florez, Steven A. Lubitz, Daniel G. MacArthur, James B. Meigs, Benjamin M. Neale, Aarno Palotie, Samuli Ripatti, Jeremiah Scharf, James S. Ware, Benjamin M. Neale, Mark J. Daly & Daniel G. MacArthur
2. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

Sanna Gudmundsson, Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alfoldi, Qingbo Wang, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Matthew Solomonson, Nicholas A. Watts, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Christopher Vittal, Irina M. Armean, Mark J. Daly, Daniel G. MacArthur, Benjamin M. Neale, Benjamin M. Neale, Mark J. Daly & Daniel G. MacArthur
3. Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA

Beryl B. Cummings
4. Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA

Qingbo Wang, Ryan L. Collins & Jack A. Kosmicki

5. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Ryan L. Collins, Harrison Brand, Michael E. Talkowski, Sekar Kathiresan & Jeremiah Scharf

6. Institute for Molecular Medicine Finland, Helsinki, Finland

Andrea Ganna, Leif Groop, Jaakko Kaprio, Aarno Palotie, Samuli Ripatti, Tuomi Tiinamaija, Maija Wessman & Mark J. Daly

7. Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Laura D. Gauthier, Yossi Farjoun, Eric Banks, Ben Weisburd, Louis Bergelson, Kristian Cibulskis, Miguel Covarrubias, Jeff Gentry, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade & Jose Florez

8. Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London and Barts Health NHS Trust, London, UK

Daniel Rhodes

9. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Raymond K. Walters, Katherine Tashman, Timothy Poterba, Arcturus Wang, Cotton Seed, Christopher Vittal, Mark J. Daly, Steven McCarroll, Aarno Palotie, Jeremiah Scharf, Benjamin M. Neale & Mark J. Daly

10. National Heart & Lung Institute and MRC London Institute of Medical Sciences, Imperial College London, London, UK

Nicola Whiffin, James S. Ware, Jaspal Kooner & James S. Ware

11. Cardiovascular Research Centre, Royal Brompton & Harefield Hospitals NHS Trust, London, UK

Nicola Whiffin, James S. Ware & James S. Ware

12. Department of Pediatrics, University of Washington, Seattle, WA, USA

Jessica X. Chong

13. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

Kaitlin E. Samocha

14. Vertex Pharmaceuticals Inc, Boston, MA, USA

Zachary Zappala

15. Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

Sanna Gudmundsson & Anne H. O'Donnell-Luria

16. Department of Pediatrics, Harvard Medical School, Boston, MA, USA

Anne H. O'Donnell-Luria

17. Department of Genetics, Yale School of Medicine, New Haven, CT, USA

Monkol Lek

18. Broad Genomics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Kristen M. Connolly, Steven Ferriera, Stacey Gabriel & Namrata Gupta

19. Department of Neurology, Harvard Medical School, Boston, MA, USA

Michael E. Talkowski

20. Unidad de Investigacion de Enfermedades Metabolicas, Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico

Carlos A. Aguilar Salinas

21. Peninsula College of Medicine and Dentistry, Exeter, UK

Tariq Ahmad

22. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA

Christine M. Albert & Daniel Chasman

23. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Christine M. Albert

24. Department of Cardiology, University Hospital, Parma, Italy

Diego Ardissino

25. Department of Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel

Gil Atzmon

26. Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA

Gil Atzmon

27. Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, USA

Gil Atzmon

28. Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

John Barnard & Mina K. Chung

29. Sorbonne Université, APHP, Gastroenterology Department, Saint Antoine Hospital, Paris, France

Laurent Beaugerie & Harry Sokol

30. Framingham Heart Study, National Heart, Lung, & Blood Institute and Boston University, Framingham, MA, USA

Emelia J. Benjamin & Josée Dupuis

31. Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Emelia J. Benjamin

32. Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

Emelia J. Benjamin

33. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

Michael Boehnke

34. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Lori L. Bonnycastle

35. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Erwin P. Bottinger, Judy Cho & Ruth J. F. Loos

36. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA

Donald W. Bowden

37. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

Donald W. Bowden

38. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

Donald W. Bowden

39. Department of Cardiovascular Sciences and NIHR Leicester Biomedical Research Centre, University of Leicester, Leicester, UK

Matthew J. Bown & Nilesh J. Samani

40. NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK

Matthew J. Bown & Nilesh J. Samani

41. Department of Epidemiology and Biostatistics, Imperial College London, London, UK

John C. Chambers & Jaspal Kooner

42. Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK

John C. Chambers & Jaspal Kooner

43. Imperial College Healthcare NHS Trust, Imperial College London, London, UK

John C. Chambers & Jaspal Kooner

44. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China

Juliana C. Chan & Ronald C. W. Ma

45. Department of Medicine, Harvard Medical School, Boston, MA, USA

Daniel Chasman, Bruce Cohen, Gad Getz, Sekar Kathiresan, James B. Meigs & Dost Ongur

46. Program for Neuropsychiatric Research, McLean Hospital, Belmont, MA, USA

Bruce Cohen & Dost Ongur

47. Department of Medicine, University of Mississippi Medical Center, Jackson, MI, USA

Adolfo Correa

48. Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA

Dana Dabelea

49. Department of Medicine and Pharmacology, University of Illinois at Chicago, Chicago, IL, USA

Dawood Darbar

50. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

Ravindranath Duggirala

51. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

Josée Dupuis

52. Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

Patrick T. Ellinor & Steven A. Lubitz

53. Cardiovascular Epidemiology and Genetics, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Catalonia, Spain

Roberto Elosua

54. Centro de Investigación Biomédica en Red Enfermedades Cardiovasculares (CIBERCV), Barcelona, Catalonia, Spain

Roberto Elosua & Jaume Marrugat

55. Department of Medicine, Medical School, University of Vic-Central University of Catalonia, Vic, Catalonia, Spain

Roberto Elosua

56. Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany

Jeanette Erdmann

57. DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck, Germany

Jeanette Erdmann

58. University Heart Center Lübeck, Lübeck, Germany

Jeanette Erdmann

59. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

Tõnu Esko & Andres Metspalu

60. Helsinki University and Helsinki University Hospital, Clinic of Gastroenterology, Helsinki, Finland

Martti Färkkilä

61. Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

Jose Florez

62. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Jose Florez & Pamela Sklar

63. Program in Metabolism, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Jose Florez

64. Institute of Clinical Molecular Biology (IKMB), Christian-Albrechts-University of Kiel, Kiel, Germany

Andre Franke

65. Bioinformatics Consortium, Massachusetts General Hospital, Boston, MA, USA

Gad Getz

66. Cancer Genome Computational Analysis Group, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Gad Getz

67. Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

Gad Getz

68. Cancer Center, Massachusetts General Hospital, Boston, MA, USA

Gad Getz

69. Endocrinology and Metabolism Department, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

Benjamin Glaser

70. Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY, USA

Stephen J. Glatt

71. Institute for Genomic Medicine, Columbia University Medical Center, Hammer Health Sciences, New York, NY, USA

David Goldstein

72. Department of Genetics and Development, Columbia University Medical Center, Hammer Health Sciences, New York, NY, USA

David Goldstein

73. Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Cuernavaca, Mexico

Clicerio Gonzalez

74. Genomics, Diabetes and Endocrinology, Lund University, Lund, Sweden

Leif Groop

75. Lund University Diabetes Centre, Malmö, Sweden

Christopher Haiman

76. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA

Craig Hanis

77. Department of Neurology, Columbia University, New York, NY, USA

Matthew Harms

78. Institute of Genomic Medicine, Columbia University, New York, NY, USA

Matthew Harms

79. Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland

Mikko Hiltunen

80. Department of Psychiatry, Helsinki University Central Hospital, Lapinlahdentie, Helsinki, Finland

Matti M. Holi

81. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Christina M. Hultman & Patrick F. Sullivan

82. Icahn School of Medicine at Mount Sinai, New York, NY, USA

Christina M. Hultman

83. Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland

Mikko Kallela

84. Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Jaakko Kaprio, Samuli Ripatti & Erkki Vartiainen

85. Cardiovascular Disease Initiative and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Sekar Kathiresan

86. Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, South Korea

Bong-Jo Kim & Young Jin Kim

87. MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University School of Medicine, Cardiff, UK

George Kirov, Michael C. O'Donovan & Michael J. Owen

88. Department of Health, National Institute for Health and Welfare (THL), Helsinki, Finland

Seppo Koskinen

89. Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

Harlan M. Krumholz

90. Division of Pediatric Gastroenterology, Emory University School of Medicine, Atlanta, GA, USA

Subra Kugathasan

91. Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea

Soo Heon Kwak & Kyong Soo Park

92. Institute of Clinical Medicine, The University of Eastern Finland, Kuopio, Finland

Markku Laakso

93. Kuopio University Hospital, Kuopio, Finland

Markku Laakso

94. Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

Terho Lehtimäki & Kari M. Mattila

95. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Ruth J. F. Loos

96. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China

Ronald C. W. Ma

97. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China

Ronald C. W. Ma

98. Cardiovascular Research REGICOR Group, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Catalonia, Spain

Jaume Marrugat

99. Department of Genetics, Harvard Medical School, Boston, MA, USA

Steven McCarroll

100. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Headington, Oxford, UK

Mark I. McCarthy

101. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

Mark I. McCarthy

102. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK

Mark I. McCarthy

103. F Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Dermot McGovern

104. Atherogenomics Laboratory, University of Ottawa Heart Institute, Ottawa, Canada

Ruth McPherson

105. Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA

James B. Meigs

106. Department of Clinical Sciences, University Hospital Malmo Clinical Research Center, Lund University, Malmo, Sweden

Olle Melander

107. Department of Clinical Sciences, Lund University, Skane University Hospital, Malmo, Sweden

Peter M. Nilsson

108. Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico

Lorena Orozco

109. Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK

Colin N. A. Palmer

110. Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea

Kyong Soo Park

111. Department of Psychiatry, Keck School of Medicine at the University of Southern California, Los Angeles, CA, USA

Carlos Pato

112. Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Ann E. Pulver

113. Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK

Nazneen Rahman

114. Medical Research Center, Oulu University Hospital, Oulu, Finland and Research Unit of Clinical Neuroscience, Neurology, University of Oulu, Oulu, Finland

Anne M. Remes

115. Research Center, Montreal Heart Institute, Montreal, Quebec, Canada

John D. Rioux

116. Department of Medicine, Faculty of Medicine, Université de Montréal, Quebec, Canada

John D. Rioux

117. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

Dan M. Roden

118. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

Dan M. Roden

119. Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Danish Saleheen

120. Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Danish Saleheen

121. Center for Non-Communicable Diseases, Karachi, Pakistan

Danish Saleheen

122. National Institute for Health and Welfare, Helsinki, Finland

Veikko Salomaa & Jaana Suvisaari

123. Deutsches Herzzentrum München, Munich, Germany

Heribert Schunkert

124. Technische Universität München, Munich, Germany

Heribert Schunkert

125. Division of Cardiovascular Medicine, Nashville VA Medical Center and Vanderbilt University, School of Medicine, Nashville, TN, USA

Moore B. Shoemaker

126. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Pamela Sklar

127. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Pamela Sklar

128. Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Pamela Sklar

129. Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland

Hilkka Soininen

130. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

Tim Spector

131. Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA

Patrick F. Sullivan

132. Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore, Singapore

E. Shyong Tai & Yik Ying Teo

133. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

E. Shyong Tai

134. Duke-NUS Graduate Medical School, Singapore, Singapore

E. Shyong Tai

135. Life Sciences Institute, National University of Singapore, Singapore, Singapore

Yik Ying Teo

136. Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

Yik Ying Teo

137. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland

Tuomi Tiinamaija & Maija Wessman

138. HUCH Abdominal Center, Helsinki University Hospital, Helsinki, Finland

Tuomi Tiinamaija

139. Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego, CA, USA

Ming Tsuang

140. Institute of Genomic Medicine, University of California, San Diego, CA, USA

Ming Tsuang

141. Juliet Keidan Institute of Pediatric Gastroenterology, Shaare Zedek Medical Center, The Hebrew University of Jerusalem, Jerusalem, Israel

Dan Turner

142. Instituto de Investigaciones Biomédicas UNAM, Mexico City, Mexico

Teresa Tusie-Luna

143. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico

Teresa Tusie-Luna

144. Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Hugh Watkins

145. Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands

Rinse K. Weersma

146. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA

James G. Wilson

147. Program in Infectious Disease and Microbiome, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Ramnik J. Xavier

148. Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA

Ramnik J. Xavier

149. Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, CA, USA

Marquis P. Vawter

Consortia

Genome Aggregation Database Consortium

- , Carlos A. Aguilar Salinas
- , Tariq Ahmad
- , Christine M. Albert
- , Diego Ardiissino
- , Gil Atzmon
- , John Barnard
- , Laurent Beaugerie
- , Emelia J. Benjamin
- , Michael Boehnke
- , Lori L. Bonnycastle
- , Erwin P. Bottinger
- , Donald W. Bowden
- , Matthew J. Bown
- , John C. Chambers
- , Juliana C. Chan
- , Daniel Chasman
- , Judy Cho
- , Mina K. Chung
- , Bruce Cohen
- , Adolfo Correa
- , Dana Dabelea
- , Mark J. Daly
- , Dawood Darbar
- , Ravindranath Duggirala

- , Josée Dupuis
- , Patrick T. Ellinor
- , Roberto Elosua
- , Jeanette Erdmann
- , Tõnu Esko
- , Martti Färkkilä
- , Jose Florez
- , Andre Franke
- , Gad Getz
- , Benjamin Glaser
- , Stephen J. Glatt
- , David Goldstein
- , Clicerio Gonzalez
- , Leif Groop
- , Christopher Haiman
- , Craig Hanis
- , Matthew Harms
- , Mikko Hiltunen
- , Matti M. Holi
- , Christina M. Hultman
- , Mikko Kallela
- , Jaakko Kaprio
- , Sekar Kathiresan
- , Bong-Jo Kim
- , Young Jin Kim
- , George Kirov
- , Jaspal Kooner
- , Seppo Koskinen
- , Harlan M. Krumholz
- , Subra Kugathasan
- , Soo Heon Kwak
- , Markku Laakso
- , Terho Lehtimäki
- , Ruth J. F. Loos
- , Steven A. Lubitz
- , Ronald C. W. Ma
- , Daniel G. MacArthur
- , Jaume Marrugat
- , Kari M. Mattila
- , Steven McCarroll
- , Mark I. McCarthy
- , Dermot McGovern

- , Ruth McPherson
- , James B. Meigs
- , Olle Melander
- , Andres Metspalu
- , Benjamin M. Neale
- , Peter M. Nilsson
- , Michael C. O'Donovan
- , Dost Ongur
- , Lorena Orozco
- , Michael J. Owen
- , Colin N. A. Palmer
- , Aarno Palotie
- , Kyong Soo Park
- , Carlos Pato
- , Ann E. Pulver
- , Nazneen Rahman
- , Anne M. Remes
- , John D. Rioux
- , Samuli Ripatti
- , Dan M. Roden
- , Danish Saleheen
- , Veikko Salomaa
- , Nilesh J. Samani
- , Jeremiah Scharf
- , Heribert Schunkert
- , Moore B. Shoemaker
- , Pamela Sklar
- , Hilkka Soininen
- , Harry Sokol
- , Tim Spector
- , Patrick F. Sullivan
- , Jaana Suvisaari
- , E. Shyong Tai
- , Yik Ying Teo
- , Tuomi Tiinamaija
- , Ming Tsuang
- , Dan Turner
- , Teresa Tusie-Luna
- , Erkki Vartiainen
- , Marquis P. Vawter
- , James S. Ware
- , Hugh Watkins

- , Rinse K. Weersma
- , Maija Wessman
- , James G. Wilson
- & Ramnik J. Xavier

Corresponding authors

Correspondence to [Konrad J. Karczewski](#) or [Daniel G. MacArthur](#).

Supplementary information

Supplementary Table

Homozygous pLoF variants in genes associated with autosomal recessive and X-linked phenotypes in OMIM.

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

Gudmundsson, S., Karczewski, K.J., Francioli, L.C. *et al.* Addendum: The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **597**, E3–E4 (2021). <https://doi.org/10.1038/s41586-021-03758-y>

- Published: 09 August 2021

- Issue Date: 02 September 2021
 - DOI: <https://doi.org/10.1038/s41586-021-03758-y>
-

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03758-y>.

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